DOCUMENT RESUME

ED 407 444 TM 027 116

AUTHOR	Hatry, Harry P.; Kopczynski, Mary
TITLE	Guide to Program Outcome Measurement for the U.S. Department of Education.
INSTITUTION	Urban Inst., Washington, D.C.
SPONS AGENCY	Department of Education, Washington, DC. Planning and Evaluation Service.
PUB DATE	Feb 97
NOTE	124p.; Prepared for the Planning and Evaluation Service by the Urban Institute under a subcontract from the Research Triangle Institute, RTI Subcontract No. 8-36U-5439.
CONTRACT	LC92008001
AVAILABLE FROM	Planning and Evaluation Service, Publications Office, U.S. Department of Education, Room 4163 FB10B, Washington, DC 20202; fax: 202-401-3036.
PUB TYPE	Reports - Evaluative (142)
EDRS PRICE	MF01/PC05 Plus Postage.
DESCRIPTORS	Educational Objectives; *Educational Research; Elementary Secondary Education; *Evaluation Methods; Evaluation Utilization; Higher Education; *Measurement Techniques; *Outcomes of Education; Pilot Projects; Program Development; Program Evaluation; *User Needs (Information)
IDENTIFIERS	*Department of Education; Educational Indicators; Monitoring

ABSTRACT

This volume is intended as a guide to managers of U.S. Department of Education programs and their staffs to help them in their efforts to design reasonably valid, reliable, and useful outcome monitoring procedures. It provides recommendations for development of an outcome measurement process for individual educational programs. These suggestions can be used for a program that has not yet developed an evaluation process or to improve an existing process. After a discussion of preliminary steps, the guide suggests the following steps to developing the outcome measurement system: (1) identify the program's mission and objectives and its customers; (2) identify the outcomes that should be monitored; (3) select outcome indicators; (4) identify data sources and data collection procedures; (5) select outcome indicator breakouts; (6) compare the findings to benchmarks; (7) pilot test the procedures; (8) analyze and report outcome information; and (9) use outcome information. Key issues in these processes are summarized. It must be recognized that unless the measurement system produces information that is useful to the program, the effort will have been wasted. Four appendixes present sample teacher and student surveys and program outcome indicators from the Star Schools program, as well as a discussion of trained observer procedures. (Contains 37 exhibits and 21 references.) (SLD)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

TM

ED 407 444

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION
CENTER (ERIC)
✓ This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.
• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

GUIDE TO PROGRAM OUTCOME MEASUREMENT

for the U.S. Department of Education

Working Document
February 1997

Harry P. Hatry and Mary Kopczynski

The Urban Institute
2100 M Street, N.W.
Washington, D.C.

TM027116

Prepared for the Planning and Evaluation Service of the U.S. Department of Education

BEST COPY AVAILABLE

GUIDE TO PROGRAM OUTCOME MEASUREMENT

for the U.S. Department of Education

Harry P. Hatry and Mary Kopczynski

The Urban Institute
2100 M Street, N.W.
Washington, D.C.

February 1997

Prepared for the Planning and Evaluation Service, U.S. Department of Education by the Urban Institute under a subcontract from the Research Triangle Institute (ED Contract No. LC92008001, RTI Subcontract No. 8-36U-5439). The material contained in this manual represents the work of the Urban Institute and the opinions expressed do not necessarily reflect the opinions or the policies of RTI or the Department.

To obtain additional copies of this manual contact:

Geneise Cooke
Planning and Evaluation Service
Publications Office
U.S. Department of Education
Room 4163 FB10B
Washington, D.C. 20202

Phone: (202) 401-0590
Fax: (202) 401-3036
E-mail: geneise_cooke@ed.gov

To send comments on improvements to the manual or ways for the Department to support outcome measurement, contact:

Ann Nawaz
Planning and Evaluation Service
U.S. Department of Education
Washington, DC 20202-8170

Phone: (202) 401-5344
Fax: (202) 401-3036
E-mail: ann_nawaz@ed.gov

Contents

Acknowledgments

Section 1: Overview 1
Introduction 1
Terminology 3
Performance and Outcome Measurement 3
Categories of Performance Information 3
Additional Comments and Discussion 7

Section 2: Getting Started - Preliminary Steps 11
Determine the Programmatic Scope to Be Included 11
Secure Top-Level Office Support for the Outcome Measurement Effort 11
Establish a Working Group to Oversee Development of the Outcome Measurement Process 12

Section 3: Developing the Outcome Measurement System 17

Step 1. Identify the Program's Mission/Objectives and Customers 17
Mission/Objective Statement 17
Identifying Categories of Customers 19

Step 2. Identify the Outcomes that Should Be Monitored 21
Sources of Information on Program Outcomes 21
Focus Groups 21
Meetings and Related Input From State and Local Personnel or Other "Partners" 23
Outcome-Sequence Charts/Logic Models 24
Role-Playing By Program Staff 26
Gathering Candidate Outcomes Identified From All Sources 29

Step 3. Select Outcome Indicators 31
Use of Outcome Sequence Charts 34
Tips on Selecting Outcome Indicators 34
Options for Statistical Forms of the Indicators 35
A Useful Option: Identify the Extent of Program's Influence/Control Over Each Outcome Indicator 35

Step 4. Identify Data Sources and Data Collection Procedures 39
Major Sources of Outcome Data 39
Outcome Information Obtained from Program and Agency Records 39
Administered Tests 40
Customer Surveys 41
Trained Observer Procedures 49
Identifying Data Collection Procedures 50

Step 5. Select Outcome Indicator Breakouts 53
Transforming the Data into Really Useful Information 53
Categories of Breakouts for Outcome Data 53
Geographical Breakouts 53
Organizational Unit/Project Breakouts 54
Customer Characteristics Breakouts 54
Degree-of-Difficulty Breakouts 55
Service Delivery Process or Procedure Breakouts 57
Procedures for Selecting, and Then Collecting Data on Breakouts 59

Step 6. Compare Findings to Benchmarks 61
Categories of Benchmarks 61
Comparisons of Current to Previous Performance 61
Comparisons of the Performance of Similar Units 61
Comparisons of Outcomes for Different Customer Groups 62
Comparisons to Pre-Selected Targets 62
Comparisons of Different Service Delivery Practices 64

Step 7. Pilot Test the Procedures 67

Step 8. Analyze and Report Outcome Information 69
Major Analysis Elements 69
Reporting 70
Provide Explanatory Information For Both Analysis and Reporting 72
Disseminate the Outcome Reports 73
Document the Outcome Measurement Process 73

Step 9. Use Outcome Information 83
What Outcome Data Are Not Likely To Tell You 83
Suggestions For Making Outcome Information More Useful 83
Uses for Program Outcome Information 84
Special Note on Use of Outcome Information for Grants and Contracts 85
Quality Control 85
Some Special Problems 85

Summary of Key Issues 87

References 93

Appendix 1: Sample Star Schools Program Teacher Survey 97

Appendix 2: Sample Star Schools Program Student Survey 101

Appendix 3: Proposed Outcome Indicators for Star Schools Program Annual Performance Measurement 105

Appendix 4: Trained Observer Ratings 111

List of Exhibits

1. Categories of Performance Information 4
2. Typical Service Quality Characteristics 6
3. Developing an Outcome Measurement Process 10
4. Topics to be Addressed by Outcome Measurement Working Groups 14
5. Sample Working Group Meeting Topics 15
6. Sample Project Schedule 16
7. Example of a Mission/Objective Statement Distance Learning Programs 18
8. Sources of Information on Program Mission/Objectives and Customers 19
9. Introduction to a Focus Group 22
10. Parent Involvement/Drop-Out Prevention: Sample Outcome Sequence Chart 25
11. Systemic Reform: Illustrative Conceptual Outcome Sequence Chart 27
12. Star Schools: Sample Outcome Sequence Chart 28
13. Some Criteria for Selecting Outcome Indicators 31
14. A Checklist for Outcome Indicators 32
15. Illustrative Performance Indicators: Parent Resource Centers 33
16. Systemic Reform: Conceptual Outcome Sequence Chart Using Illustrative Performance Indicators 36
17. Star Schools: Outcome Sequence Chart Using Illustrative Performance Indicators 37
18. Selected Outcome Indicators for External Reporting Purposes: Star Schools Program 38
19. Information Obtainable from Customer Surveys 41
20. Suggestions for Increasing Response Rates to Mailed Customer Surveys 45
21. Sample Transmittal Letter 45
22. Elements that Should be Included in a Contract for a Customer Survey 48
23. Categories of Breakouts for Outcome Data 54
24. Examples of Customer Characteristics by Which Outcome Indicators Should be Broken Out (Disaggregated) 55
25. Workload (Client) Difficulty Breakout 56

26. Comparing Program Procedures for Processing Loan Applications: Computer vs. Manual Procedures 58
27. Example of Comparison of Outcomes Before vs. After Procedure Changes 65
28. Reporting Format #1: Actual Outcomes vs. Targets Using Indicators Developed by the Star School Working Group (Illustration Only) 74
29. Reporting Format #2: Comparisons Across Geographical Locations (Illustrations Only) 75
30. Reporting Format #3: Comparisons of Actual Values vs. Target — Current and Cumulative: Parental Resource Centers (Illustrative Only) 75
31. Reporting Format #4: Outcomes by Individual Project, by Achievement Level (Illustration Only) 76
32. Reporting Format #5: Breakouts of Responses to a Customer Survey by Demographic or Program Characteristics 77
33. Parent Participation in Specific School Activities 78
34. High School Dropouts Multomah County, 1990-1995 79
35. Sample School Report Card for Parents, Taxpayers, and the Business Community 80
36. Outcome Indicator Specification Sheet 81
37. Trained Observer Rating Scale: "Condition of Classroom Walls" 114

Foreword

Department program managers, and their staffs, should routinely track the outcomes, the results, of their programs. This is just common sense. Information on outcomes should be key information to programs to help guide them on program improvement needs. In past decades obtaining outcome information has been very difficult. We have lacked the data collection and data processing tools. However, now in the mid 1990s, these are no longer major obstacles.

Furthermore the Government Performance and Results Act of 1993, passed unanimously by both Houses of Congress, *requires* that Department programs provide annual plans that include program outcome indicators. Target values for the coming fiscal year for each indicator are also required. The plans are required to be submitted to OMB with the FY 1999 budget request in fall 1997.

The Department of Education seeks to work with, and assist, states and local school districts to pursue excellence in education for all students. No less should the Department strive to excellence in our own activities. This requires us to track, analyze, and report regularly on our progress in achieving outcomes.

At the same time, we all need to recognize that outcome information, as pointed out in this volume, does not tell why the identified outcomes are what they are. Programs need to examine more deeply the "whys" so that specific improvement actions can be taken. At the end of later outcome reporting periods, the findings should again be examined to assess whether the hoped for improvements have occurred.

This volume is intended as a guide to program managers and their staffs to help them in their efforts to design reasonably valid, reliable, and useful outcome monitoring procedures. The material provides many detailed suggestions that both small and large Department programs should find useful. This material will be most useful to programs that are in the early stages of developing their outcome measurement systems. Even those programs that already are well along in such development, however, might find ideas here well worth using.

My office welcomes suggestions from you as to ways to help in your outcome measurement development efforts, both suggestions as to improvements to this manual and other ways that the Department can help your program to develop useful outcome measurement procedures.

Alan Ginsburg
Director, Planning and Evaluation Service
U.S. Department of Education

Acknowledgments

The authors greatly appreciate the suggestions provided by the following persons in their reviews of various drafts of this manual: Cheryl Garnette, Director of the Star Schools Program of OERI; Val Plisko, Director for of the Elementary and Secondary Education Division of the Planning and Evaluation Service; Cindy Gruhzit of Budget Service; and Steve Zwillinger of the Planning and Evaluation Service.

In addition, we received very helpful materials from Patricia Gore of Goals 2000 in OESE (especially those relating to the Parental Assistance Program) and Cheryl Garnette and her staff at the Star Schools Program. We are also grateful to Dr. Becky Hayward of the Research Triangle Institute for her support throughout the work. Susan Thompson-Hoffman of Planning and Evaluation Service provided considerable ideas for our work, especially in its early stages. Finally, we gratefully acknowledge the considerable help and support throughout our work from Nancy Rhett of Planning and Evaluation, Service, the Department project manager for this project. She also provided many useful suggestions on drafts of this manual.

Section 1: Overview

Introduction

The primary new thrust of the GPRA legislation, of the Executive Order on Setting Customer Service Standards (E.O. 12862), and of the Department's own emerging performance measurement process is on indicators of outcomes. This guide provides recommendations for development of an outcome measurement process for individual educational programs. The relationship to these other types of performance indicators has been discussed, but this guide focuses on outcome indicators.

Developing program outcome information is a critical step in developing a high quality performance measurement system for programs and projects in the federal government. Good program management requires collection and use of outcome data to provide guidance for improvement. Without information on results, managers can only supervise "inputs" or monitor processes. Decisions on whether the program is actually working well—or what needs changing—are made in the absence of hard data on actual outcomes.

Two recent national initiatives have reinforced the importance of developing performance information on program results:

- The Government Performance and Results Act of 1993 (GPRA) strongly reinforced the importance of managing programs based on results. GPRA requires federal agencies to develop and submit an agency strategic plan and annual performance plans for its programs. Outcome performance measurements are a key part of GPRA annual performance plans.

- Vice-President Gore's National Performance Review seeks to change the culture of government through introducing modern business practices built around quality service. Reinvention involves infusing business practices such as customer service standards and surveys, streamlining and delayering of organizational operations and structures, employee empowerment, and process reengineering.

Collectively, the two reforms involve:

- A strategic plan focused on setting and achieving clear goals.
- Quality principles that frame the goals around meeting customer needs and designing improvement strategies to strengthen agency processes critical to serving government's customers.
- Performance measurement that assesses accomplishments and feeds this information back as part of a continuous improvement process.

To support these initiatives—as well as respond to serious criticism from the General Accounting Office (GAO) on its program and agency management—the U.S. Department of Education implemented an agency strategic planning and performance management process to improve the quality of management of its programs and support implementation of several major legislative reforms achieved in the early 1990s. The Department recognized the need to improve outcome measures for many of its programs and contracted with the Urban Institute to prepare this guide for program managers and staff.

Outcome measurement has four basic uses:

- *First, and foremost, outcome information should help program managers and their staff track how their programs are doing and with that information help guide improvement efforts.* The information should for example, indicate where, when, and under what conditions, outcomes appear to be satisfactory and not satisfactory.

- *Second, outcome information can be useful for developing and justifying budgets and for formulating recommendations as to needed legislation and policy.*

- *Third, outcome information is used by the President, Congress, and Department officials in helping to achieve accountability of programs for program quality and outcomes.*

- *Fourth, outcome information can be used by the program to help communicate with and inform customers and the public at large as to the extent to which education-related progress is being made.*

This guide is intended to help program managers develop and use performance measurement systems that support all four uses of outcome data. Initially the guide was intended for managers of "small" programs in the Department—programs that might not receive formal evaluations, However, this guide evolved into one that can be used by managers of programs both large and small to develop high quality performance measurement systems or improve the ones already in place.

The suggestions provided in this guide are primarily aimed at managers and staff of programs that have not yet developed a performance measurement process—especially one with satisfactory outcome measurement elements. The guide contains considerable detail on procedures, including some that are relatively "technical." Program managers, therefore, may want to read only the less detailed information and ask members of their staff to examine this document in detail.

This guide is also aimed at helping programs that already have a performance measurement process. These programs can this volume for ideas for improving their procedures, especially if they are not satisfied with their measurement of outcomes.

> GPRA requires each major Federal program to prepare annual performance plans containing outcome indicators and targets for each indicator. The first plan is due September 1997 for FY 1999. The Act also requires these programs to provide reports containing actual outcome data after the end of each year. The first annual performance reports, covering FY 1997, are due March 2000.

Terminology

Performance and Outcome Measurement

The term ***performance measurement***, as used by the Department, refers to the regular, ongoing, measurement and reporting on important performance aspects of the Department's programs, particularly outputs and outcomes. The primary focus of this new Department effort, and this report, is on tracking the outcomes (results) of programs. However, outputs, are also briefly discussed, as they relate to outcomes. Outputs represent work completed by the program. Outcomes indicate the extent to which the outputs have led to improvements sought by the program.

An ***outcome measurement process*** is the process for selecting outcome indicators and, subsequently, *regularly obtaining and reporting data on the indicators*. Outcome indicators are needed for each significant program objective.

Outcome indicators are needed at each level of the Department—at the Secretary level, at the Principal Operating Component (POC) level, and at the program level. In addition, for cross-cutting Department initiatives (such as systemic reform) performance indicators will usually be needed that cut across programs, across POCs, and perhaps even across Federal departments. This guide focuses on outcome measurement at the program level, especially the smaller (e.g., $40M or less) Department programs.

Categories of Performance Information

Outcome information should be a major part of managing public programs. Such information enables agencies to focus on results, not just on work activity and cost.

Categories of performance information are described below. (Also see Exhibit 1.) The first two types of indicators (inputs and amount of work activity) are relatively familiar. Indicators of outcomes and impact on individual programs are much more rare. Because it is so hard to obtain true indicators of impact, outcome indicators, both intermediate and end outcomes, generally will have to be used (except when the results of in-depth, formal program evaluations are available).

Only those categories described under the label "outcome" are the subject of this manual. Nevertheless, the other categories are important to programs and should also be tracked. Most of these other categories are often already tracked by Department programs. It is important for program managers and their personnel to recognize the differences between these categories of information.

1. **Inputs.** Input data indicate the amount of resources applied, for example, the amount of funds or number of employees. When related to outcome information, the combined information will provide indicators of efficiency/productivity.

2. **Process (Workload/Activity) Indicators**. Workload/activity data indicate the amount of work either pending or in process — but not completed as of the end of the reporting period. This information is very important to program managers but such data do not measure outcomes. An exception is that in which a buildup of pending, non-completed cases at the end of reporting periods is likely to delay services to customers.

Exhibit 1.
Categories of Performance Information

- Inputs—such as dollar expenditures and employee hours.
- Process indicators — indicating the amount of workload or activity.
- Outputs—amount of work completed.
- Outcomes:
  - Intermediate outcomes—program customers or partners take actions that the program seeks that are expected to lead to improved "end outcomes," such as introducing a practice encouraged by the program. Intermediate outcomes also include service qualities that cover customer concerns on *how* a service is delivered to them (such as its timeliness, accessibility, helpfulness, and accuracy), from the customers' perspective.
  - End outcomes—*the final desired results of the program's work (including the reduction of any negative effects).* For education programs these usually include improved student learning and student preparedness for the outside world and success in it.
- Efficiency/productivity in achieving *outputs* or *outcomes*—as measured by the cost, or number of employee days, per unit of output or outcome or the unit of output or outcome per dollar.
- Impact—the extent to which the program actually *caused* an outcome (especially end outcomes). To determine program impacts, in-depth program evaluations are usually needed.

The number of pending cases might be used as a surrogate indicator for delays in service to customers. (Note, however, that probably a better indicator would be a direct indicator of the extent of delays, such as the "percent of cases in which the time between the requested service and when that service was provided exceeded 'X' days," where "X" is a service standard established by the program.)

3. **Outputs.** Output data show the quantity of work activity completed. A program's outputs are expected to *lead* to desired outcomes, but outputs do *not* by themselves tell anything about the outcomes of the work done. To help identify outcomes that should be tracked, the program should ask itself what result is expected from each of its outputs.

   Examples of outputs are the amounts of: waiver requests reviewed, grant applications approved, and loan applications processed.

4. **Outcomes.** Outcomes are not *what the program itself did*, but the *consequences of* what the program did. They provide information on events, occurrences, conditions, or changes in attitudes and behavior that indicate progress toward achievement of the mission and objectives of the program. Outcomes *happen outside the program* such as to customers (e.g., students)

or to other organizations (e.g., SEAs, LEAs, individual schools, teachers, and parents) whose behavior the program hopes to affect.

The tracking of program outcomes is the central focus of this manual.

It is usually important to distinguish "intermediate outcomes" from "end" outcomes. This will help programs distinguish mission-focused results from the intermediate steps expected (but not guaranteed) to lead to those "end" results.

a. **Intermediate Outcomes.** These are outcomes that are expected to *lead to the ends desired* but are *not* themselves "ends." Sometimes, you will find it difficult to distinguish intermediate outcomes from outputs. Outputs are things that the program and its personnel have done, not things that outside persons or organizations have done.

Some examples of intermediate outcomes are:

(1) K-12 students participate in the program.
(2) School administrators report changes/improvements in classroom teaching and learning practices, relating to Department activities such as training, technical assistance, or support for new education technology.
(3) Parents report increased knowledge or awareness of local parental resource center activities.
(4) Hours of parental education received by parents.
(5) Successful completion by teachers of sponsored professional development programs.
(6) State education agencies report increased comprehensive planning encouraged by Department programs.

A Special Type of Intermediate Outcome: Program Quality Characteristics: As used in this manual, the word "quality" indicates how well a service was delivered, based on characteristics important to customers. Quality does not tell what results occurred after the service was delivered. Such characteristics are almost always important to program customers, even though the characteristics do not really represent final results. Exhibit 2 is a list of such quality characteristics that you should consider in developing your list of outcomes to track.

Tracking of these service quality outcomes should also satisfy the requirements of Presidential Executive Order 12862, "Setting Customer Service Standards." That Executive Order requires Federal agencies that provide "significant services directly to the public" to survey customers to determine "their level of satisfaction with existing services."

Some programs may choose to consider customer satisfaction with the *results* of a service as an end outcome, for example, parent satisfaction with their children's learning progress. However, customer satisfaction with characteristics of *how* a service was delivered (such as its timeliness and courteousness) and not its results should be considered an intermediate outcome.

Exhibit 2.
Typical Service Quality Characteristics

- Timeliness with which the service is provided.
- Accessibility/convenience of the service.
  - Convenience of location.
  - Convenience of hours of operations.
  - Staff availability when the customer needs the service.
- Accuracy of the assistance such as accuracy in processing customer requests for service.
- Courteousness with which the service is delivered.
- Adequacy of information disseminated to potential users about what the service is and how to obtain it.
- Condition and safety of public facilities used by customers.
- Customer satisfaction with the service.
- Customer satisfaction with the results of the service.

b. **End Outcomes.** These are the desired results of the program. End outcomes in education usually relate to effects on people, generally students or the general population. An exception might be default rates for student loans as a major outcome of a loan program—especially if defaults mean more taxpayer dollars or that fewer students can receive loans in the future. Some examples of end outcomes are:

(1) Improved student learning
(2) Improved student interest in learning
(3) Improved post-education employment, or
(4) Less disruptive behavior or violence in schools. (Some readers may prefer to consider this to be an intermediate outcome needed to achieve improved learning.)

Many programs can lead to both short-term and long-term end outcomes. For example, intervention activities and school-to-work programs have both short- and long-term ends. Increased test scores, increased school completion rates, increased skills/readiness for skilled occupations, and reduced incidents of disruption in schools can be considered as at least short-term end outcomes for many programs. Employment, ability to support a family, and reductions in welfare dependency are longer term outcomes (especially for programs aimed at students in the lower grades).

For the purposes of regular outcome monitoring, such long-term end outcomes as post-education employment and earnings are not likely to inform current program personnel on the outcomes of their current activities. Shorter term end outcomes are needed. Therefore, the shorter-term end outcomes relating to attendance, grades, behavioral problems and drop out rates are likely to be of key concern to managers and staff of such educational programs.

To do outcome *monitoring*, many programs will likely need to focus on shorter term end-outcomes than ideally desirable. To identify long-term end outcomes (such as improved rates of employment for non-college bound students and higher long-term earnings) special studies will likely be needed.

5. **Efficiency and Productivity.** These categories indicate the relation of the amount of input to the amount of output (or outcome).

   Traditionally, the ratio of the amount of input to the amount of output (or outcome) is labeled "efficiency." If you flip this ratio over to get the ratio of the amount of output (or outcome) to the amount of input, this is labeled "productivity." These are equivalent numbers.

   This manual does not focus on efficiency and productivity.

   However, one important use of outcome information, especially data on end outcomes, is to relate it to the amount of input. Use of outcome rather than output data provide a much truer picture of efficiency and productivity. (A major danger of focusing on output-to-input ratios is the temptation to increase output at the expense of results.)

   Examples of outcome-based productivity indicators are:

   - Number of school buildings that were improved from "poor" to "good" condition *per* dollar (or per employee hour);

   - Number of customers who reported that the service received had been of significant help to them *per* dollar cost of that service (or per employee hour).

   Flip these ratios over and they are called efficiency indicators. For example, if 160 customers reported being significantly helped, and the program cost $96,000:

   A. Productivity = 160/$96,000 = 1.67 customer helped per thousand dollars.
   B. Efficiency = $96,000/160 = $600 per helped customer.

Additional Comments and Discussion

What Are Some of the Distinctions Between Intermediate and End Outcomes?

Intermediate outcomes usually, but not always, occur sooner than end outcomes. Thus, intermediate outcomes are likely to provide particularly timely information for program managers. Some end outcomes occur years after a program's activities have been administered. For example, lifetime earnings are affected by schooling preparation. For long-term end outcomes for which data may not be available for many years, the program will need to focus on shorter term ends (such as improved learning and skills) and intermediate outcomes (such as completion of courses and skill development programs).

Early occurrence of an outcome, however, does not necessarily mean that it is *not* an end outcome. For example, educational activities can produce increased early learning gains and interest in education, outcomes that many (if not most) people would agree are desirable ends of educational activities. Reduced disruptions and violence in schools also can usually be considered end outcomes and can occur in the short term.

Another plus for including intermediate outcomes is that *programs almost always have more control over intermediate outcomes than they do over end outcomes.* Federal education programs almost always assist SEAs, LEAs, non-governmental organizations, and so on, rather than directly helping students. (An exception is the direct student loan program.) Changes made by these other organizations can be directly affected by the Department's programs, and also are more easily measured by the Federal programs, than are the ultimate ends sought by these programs such as student learning and preparedness gains. The latter outcomes are affected by many other factors (such as family circumstances, motivational factors, etc.) Note, however, that Federal programs seldom, if ever, will have complete control over any outcome, whether intermediate or end. Invariably, other factors outside the control of Federal education programs will also affect the actions of these other organizations.

It will not always be clear whether a particular outcome is an intermediate or end outcome. When this occurs, check the program's mission/objective statement to see how closely related the outcome is to the statement. In any case, how the outcome is labeled is much less important than that it is included in the program's measurement process.

Are "Number of Customers Served" an Output or Outcome Indicator?

"Number of customers served" is an example of an ambiguous indicator. The number of customers served is a number commonly reported by government programs. The word "served," however, is ambiguous and makes it difficult to decide which category this represents.

Programs should define more specifically what is meant by "served."

- If it means only that a program employee "saw" the customer, this seems best labeled an output.

- If the customer received some initial benefit from the service, then it can be labeled as an intermediate outcome. "Customer served" would still not be an end outcome since it does not indicate the results of that assistance.

What Is "Participation"? An Output or Outcome Indicator?

Counts of customers participating in a program can also be ambiguous and depend on the particular situation in which used.

- If attendance is mandatory, the number participating would at best be output information.

- For programs in which (a) participation is voluntary, and (b) the program includes activities aimed at attracting customers into the program (such as parental involvement programs and professional development activities), participation can be categorized as an intermediate outcome. The ability of a program to retain participants until the activities are completed (assuming participation is voluntary) is another important intermediate outcome. This type of event is more than another output because the activity has been sufficiently attractive to the customers that they have stuck with it to completion.

Remainder of this Manual

The remaining sections of this manual describe key steps in implementing and maintaining an outcome measurement process. Section 2 describes three preliminary "organizational" steps that are needed to get started. Section 3 discusses nine key steps in developing and using the outcome measurement process and the information it provides. These steps are listed in Exhibit 3.

The quality of these products depends primarily on you. This manual *only* provides guidance and suggestions.

We do *not* attempt to define what a "program" is. This is your choice. You can choose a narrow or broad scope for the program.

Exhibit 3.
Developing an Outcome Measurement Process

Preliminary Steps

- *Determine the programmatic scope to be included.* The program manager should identify and select the program coverage to be included in the outcome measurement effort. For example, it may be desirable for outcome indicators to focus on certain key program activities.

- *Secure top-level office support for the outcome measurement effort.* This support is needed in order to obtain an adequate commitment of time and resources to properly develop, implement, and operate the outcome measurement system.

- *Establish a working group to oversee development of the outcome measurement process.* The working group should be chaired by a program manager, include a variety of people familiar with the operation of the program, and be responsible for completing the following steps.

Process Development and Implementation Steps

Step 1: *Identify the program's mission/objectives, and customers.* This will assist you in answering the question, "What is successful performance?" for your program.

Step 2: *Identify the outcomes to be monitored.* Such procedures as outcome-sequence charts, role playing, and focus groups are good ways to help identify outcomes.

Step 3: *Select outcome indicators.* Be sure to identify a sufficient number of indicators to describe fully the program's accomplishments in key strategic areas.

Step 4: *Identify data sources and data collection procedures* needed to obtain data for each outcome indicator. This includes the development of data collection instruments (such as customer questionnaires) and determination of frequency of data collection and reporting.

Step 5: *Select outcome indicator breakouts.* Disaggregation of indicators is important to provide program personnel and other audiences more useful information about the conditions under which the program seems to work well, and where it does not.

Step 6: *Compare findings to benchmarks* including comparisons to previous performance, performance of similar units or similar client groups, and pre-selected targets.

Step 7: *Pilot test and revise the procedures.* Test the indicators and make revisions as needed to improve the indicators, breakouts, data sources, data collection instruments and procedures, or other elements of the outcome measurement system.

Step 8: *Analyze and report outcome information.* Examine outcomes by grade/age level, gender, minority status, location, and type of school. Seek explanations for unusual performance.

Step 9: *Use outcome information.* Incorporate outcome information into program management practices. Report on your program's performance data to supervisors, program staff, and the public whenever you have the opportunity.

Section 2: Getting Started - Preliminary Steps

The following three steps need to occur before commencing the more technical tasks of the outcome measurement process.

Determine the Programmatic Scope to Be Included

The program manager should identify and select the program coverage to be included. Many programs, even small education programs, will likely have more than one important activity. For example, the Star Schools Program sponsors projects that introduce distance learning into school systems. It also has other projects focused on national dissemination of distance learning information. The former set of projects was the focus for the program's initial outcome measurement effort.

> *Another example: parental assistance programs take many different forms. One, or all programs might be included in an initial effort to focus on different activities such as: information dissemination, parent education, early childhood programs, etc.*

If the *mission* of your program's various projects are highly similar, even though the approaches differ considerably, it is probably better to fold them into one outcome measurement process. The end outcomes sought should be similar for the projects. The intermediate outcomes, however, are likely to differ depending on the service delivery approaches used by individual projects. To the extent that the missions are significantly different from each other, separate outcome measurement procedures will likely be needed.

Secure Top-Level Office Support for the Outcome Measurement Effort

This support is needed in order to obtain an adequate commitment of time and resources to develop, implement, and operate the outcome measurement process properly. While the primary effort in developing the outcome measurement process will probably come from the program itself, many outcome measurement elements will likely require some outside support, particularly for ongoing data collection, tabulation, and analysis. Some of these activities might come from the program's own resources, such as from contract funds. However, even here such expenditures will inevitably be reviewed by upper levels. In addition, some of these tasks are likely to require special help, such as from the Department's computer services, or coordination with other data collection activities such as those undertaken by NCES. The encouragement and support of top office management are necessary to assure that at least some potential help will be available.

The program manager should also secure an adequate length of time for implementation. Because of federal legislation (Government Performance and Results Act of 1993, GPRA, which requires each major program to have a performance measurement process focusing on outcomes), programs will likely be pressed to come up with an adequate outcome measurement process in a short time period. (Full implementation is scheduled for FY 1999, with indicators to be established by September 1997.)

Most programs will require a minimum of one to two years from start to when they produce their first set of comprehensive outcome data. Few programs are currently providing much, if any, annual outcome information. The program manager should negotiate with high-level officials an overall time frame that provides a long enough period to develop and produce needed outcome information.

Many programs should be able, initially, to provide some available data on a few intermediate outcomes (for which data are more likely to be readily available than end outcome data). The program can also report that new outcome-oriented outcome indicators are in the process of being developed. This is likely to reduce the pressure on the program to have a complete outcome measurement system in place in too short and amount of time.

Establish a Working Group to Oversee Development of the Outcome Measurement Process

The manager of the program should form a working group that will oversee development of the outcome measurement process. The working group should consist of such persons as the following:

- The program manager (who probably should act as the working group facilitator);
- Members of the program staff (assuming the program is not a one-person operation);
- Representatives from related program areas in the Department;
- A representative from the relevant Office of the Assistant Secretary;
- A "technical expert," perhaps from the Office of the Under Secretary's Planning and Evaluation Service, Office of Educational Research and Improvement, or perhaps an outside consultant or contractor (preferably someone with familiarity with the Department's GPRA efforts); and
- A representative from Budget Service.

Working groups should probably be no larger than about 8-12 people. For very small programs with a one-or-two person staff, the working group could be quite small.

The working group should initially meet frequently and regularly. (Frequency and timing to some extent will depend on pressure from the Office of the Assistant Secretary and the Department to move ahead on GPRA.) The working group should plan on being in existence for one to two years (preferably two years) to work through development, implementation, and quality checking of the products of the outcome measurement process. The working group needs to address several topics. These are detailed in Exhibit 4. Exhibit 5 organizes these topics into a sample agenda for the initial working group meetings. Details on the subject matter of those meetings are discussed in later steps.

How one program organized their working group for outcome measurement:

For the Star Schools Program, the program manager chaired the group. All program staff participated (approximately seven persons). Representatives were included from the Office of the Assistant Secretary of OERI, from the Knowledge Applications Division of OERI, and from the Office of Planning and Evaluation Service. In addition, a representative from the Department's Office of Budget Services was also a member of the working group (the Department's budget person for the program). The group met several times, focusing on identification of mission, objectives, outcomes, and outcome indicators.

Before each meeting, the program manager should prepare and distribute an agenda, making the objectives of each meeting clear. In addition, the program manager should prepare a brief report on the key findings and results of the previous meeting (however, a set of detailed minutes is not likely to be needed). This report should be available for review by the group before the next meeting.

A sample schedule for developing a program outcome measurement process is presented in Exhibit 6. This schedule assumes a 15-month development process. The program, however, should provide at least annual reviews of the process for the subsequent two to three years to make sure that parts of the process are providing quality data *and* that the outcome information is being used—and is useful to the program.

Exhibit 4
Topics to be Addressed by Outcome Measurement Working Groups

1. The purpose of the working group.

2. The mission objectives and clients of the program.

3. The outcomes that the program seeks. (As discussed in Step 2, this includes both intermediate and end outcomes.)

4. Needed meetings with interest groups, such as client groups (perhaps in individual interviews or focus groups) in order to identify outcomes desired from the viewpoints of these interest groups.

5. Specific outcome indicators for measuring each outcome.

6. Appropriate data sources for each outcome indicator.

7. The specific data collection procedures needed to obtain data on the indicators, especially new data (including development of data collection instruments such as survey questionnaires).

8. The specific breakouts needed for each indicator, such as breakouts of student achievement by student demographic characteristics, location, type of approach used, etc. (Breakout information can be extremely useful in determining under what conditions successful outcomes are occurring.)

9. Planning, undertaking, and reviewing pilot test of the new data collection procedures.

10. Formats for presenting the outcome information (so as to be informative and user-friendly).

11. Determination of the roles that program partners (such as project grantees) should play in developing and implementing the performance measurement process. If the program uses a national evaluation contractor, the role of the contractor in providing annual outcome data should be considered. For example, the Star Schools Program supports a number of projects, each of which have project personnel (who provides services to school systems) and a project evaluator. Both project personnel and project evaluators participated in helping the program identify outcomes and outcome indicators. Some of them also agreed to pilot-test some of the new data collection instruments.

12. The time schedule for undertaking the above items, for pilot-testing the procedures, and for subsequently making modifications based on the pilot results. A sample project schedule is shown in Exhibit 6.

13. A long-term schedule for implementation, such as a three-year schedule, indicating the timing of data collection and analysis relevant to each year's budgeting cycle and who is responsible for what.

14. The uses for the outcome information—both to the program personnel themselves (to help improve the program), to grantees, and to the ultimate customers.

Exhibit 5
Sample Working Group Meeting Topics

Meeting One

1. Identify the purposes and uses for outcome data.
2. Discuss working group mission, objectives, and overall schedule.
3. Begin defining program mission, objectives, and customers.
4. Plan for focus groups

Meeting Two

5. Complete defining program mission, objectives, and customers.
6. Begin identifying outcomes to be tracked.
7. Role-play as customers.
8. Prepare outcome sequence charts.
9. Work out details of focus groups (to be held before Meeting Three).

Meeting Three

10. Review findings from focus groups.
11. Finalize list of candidate outcomes to track.
12. Begin identifying outcome indicators.
13. Discuss possible data sources and data collection procedures.

Meeting Four

14. Work on identifying outcome indicators, data sources, and basic collection procedures.
15. Identify desirable breakouts of indicator data.
16. Plan for development of detailed data collection procedures such as customer survey questionnaires.

Meeting Five

17. Finalize outcome indicators and data sources.
18. Review initial cuts at detailed data collection procedures such as customer survey questionnaires.
19. Begin planning for pilot testing of new data collection procedures.
20. Complete plan for pilot test and initiate it.

Meetings Seven, Eight, and Nine

21. Review progress of pilot test.
22. Work out test problems.
23. Select outcome report formats and identify needed tabulations for the outcome data coming from the pilot test.

Meeting Ten

24. Review results of pilot test procedures.
25. Identify and make necessary modifications.
26. Begin reviewing pilot test outcome data.

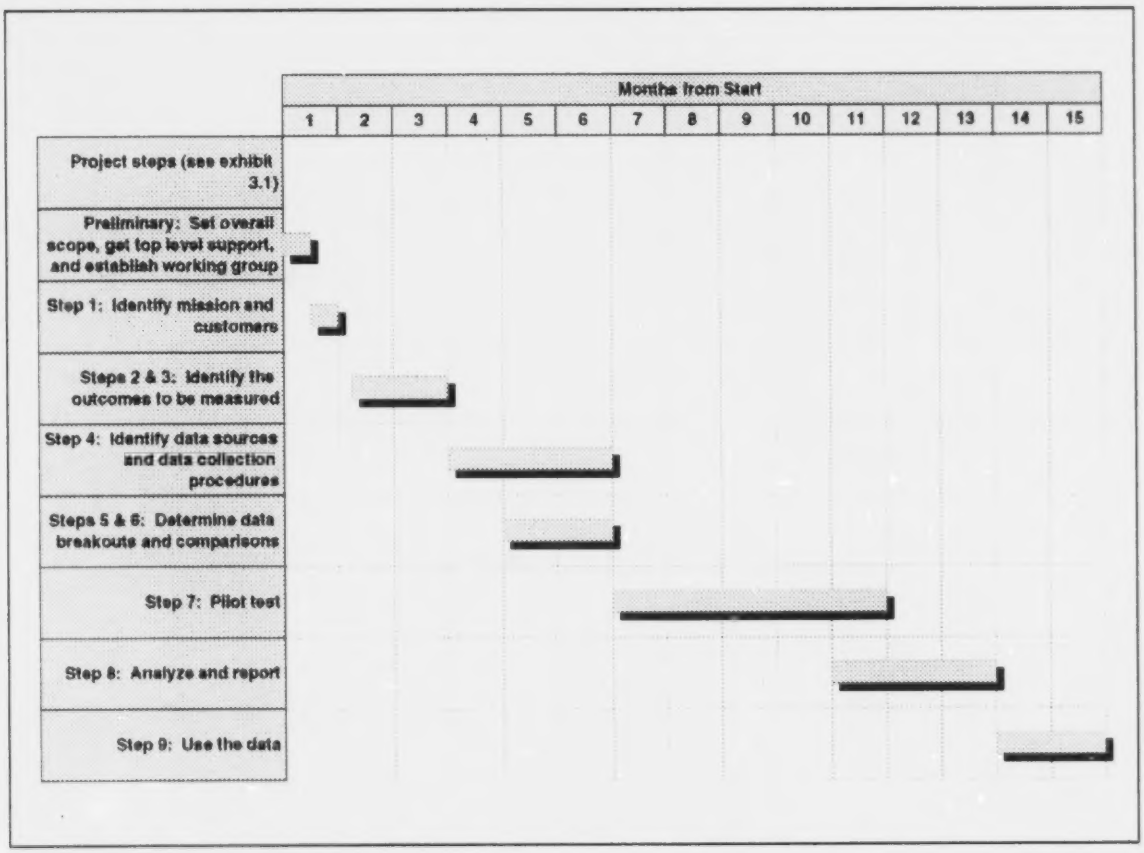
Meeting Eleven

27. Begin documenting outcome measurement procedures for ongoing implementation.
28. Identify specific ways to make the outcome data most useful (by determining frequency of reporting, methods of report dissemination, and ways to follow up on findings).

Meeting Twelve

29. Review all aspects of the outcome measurement process.
30. Finalize documentation.
31. Develop a multi-year schedule for full implementation.

Exhibit 6:
Sample Project Schedule



Section 3: Developing the Outcome Measurement System

The following nine steps detail the activities a program needs to undertake, and then use, to implement a successful outcome measurement system.

Step 1. Identify the Program's Mission/Objectives and Customers

Your first step is to prepare a mission/objective statement for the program. This tells where you want to go, i.e., what overarching results are hoped to be achieved, or the purpose toward which the program is directed.[1]

Mission/Objective Statement

This is a statement that expresses the *major results* sought by the program. If desired, you can also add a statement identifying the primary way in which the program provides its services.

This is your chance to step back and think about the mission and objectives of the program.

Here are a few tips:

1. Focus on the results sought from the program's work and activities—how program activities are hoped to affect customers and the public.
2. Most programs have multiple objectives. It is better to include too many objectives in your mission/objective statement than to eliminate ones that later may be found to be important.
3. The mission/objective statement is the starting point for identifying the outcomes to be measured and the specific performance indicators that are needed.

The basic form of a mission/objective statement is as follows:

To: [Identify here the basic objectives (results) that the program seeks. Include any major *negative* consequences that the program seeks to avoid.]

By: [Identify the basic way the service is provided. DANGER: Avoid detail and do not constrain your options on ways to provide the service. The program is more likely to be stimulated to try different approaches when it focuses on the mission/objectives.]

An example from the Star Schools Program is shown in Exhibit 7. In this example, the "To" statement includes potential intermediate outcomes such as improved instruction and student access to a wide range of subjects. The specific approach of the program is use of distance learning technologies.

[1] The word "objective" refers to a more specific set of program purposes that flow from the mission statement.

Exhibit 7.
Example of a Mission/Objective Statement
Distance Learning Programs

To: Improve student learning and employability through providing access to, and improving instruction in, a wide range of subjects.

By: The use of distance learning technologies.

The "By" statement is not necessary for those programs where the basic approach is expected to be clear to all users of the performance information.

Mission/objective statements should have the following characteristics:

- The "To" statement is a *general* statement of the major missions/purposes/objectives/results that the program would like to achieve.
- The statement should not contain *numerical* targets (such as "improve an outcome by 15 percent"). Such targets, however, should be developed separately, as discussed in Step 6.
- The statement should identify all the major objectives that the program hopes to achieve.
- If the program is likely to involve important potential *negative unintended effects*, the statement should include words explicitly calling for minimizing these effects. For example, the mission statement might include words such as "... and to minimize negative effects such as [identify the major possible negative impacts].

Examples of unintended effects are:

- Discouraging and hurting morale of non-technical teachers in schools with classes where distance learning technologies are used.
- Incurring parent (and public) opposition to a new educational approach supported by a federal program.
- Attracting better students and teachers to high-tech classes, leaving more needy students in worsened learning situations.

Sources of information to help identify the program's mission/objectives are listed in Exhibit 8.

Exhibit 8
Sources of Information on Program Mission/Objectives and Customers

- Legislation and regulations,
- Mission statements contained in budget documents,
- Strategic plans (Department, POC, or program),
- Various program descriptions and annual reports,
- Discussions with upper level officials and their staffs,
- Discussions with legislators and their staffs,
- Discussions or meetings with customers and service providers,
- Input from program personnel,
- Complaint information (What have customers complained about?),
- Mission statements used by other levels of government for similar programs.

Identifying Categories of Customers

Your mission/objective statement usually should identify who your customers are unless you believe this will be obvious to most users of the outcome information. Exhibit 7 above identifies students as the primary customers.

However, "who your customers are" may not be as obvious as it seems.

Ask such questions as the following:

- Who benefits from the program?
- Who might be hurt by program activities?

This may also help you identify potential unintended negative effects of the program that should be identified in the mission/objective statement.

- What persons not directly targeted by the program could be significantly affected by the program?

- Which particular demographic or interest groups are particularly affected by the program?

- Is the public-at-large likely to have a major interest in what the program accomplishes (rather than just what it costs)? For example, for parenting-education programs, parents might be considered a key customer group.

Some examples of key customer groups in various education programs:

- For the Star Schools program, the Working Group identified the following customers: K-12 students, teachers (who could benefit from distance learning professional development programs) parents, school administrators, adult learners, other educators, high school dropouts, and residents in correctional facilities.

- For school-to-work opportunity programs, customers would likely include not only the targeted audience of students and recent dropouts, but also prospective employers and parents.

- For a program aimed at combating teenage pregnancy, the primary customers will be teenagers. However, should males as well as females be targeted by such programs? Are parents also customers?

- For school violence prevention programs, the whole student body and school personnel are both likely to be beneficiaries.

- For many systemic reform programs, state and local educational agencies are the immediate customers, with students being the ultimate customers.

Step 2. Identify the Outcomes that Should Be Monitored

The purpose of this step is to identify the specific *outcomes* that should be monitored by the program. (The next step will be to translate each outcome into specific, measurable outcome indicators.) All relevant important outcomes should be identified.

Sources of Information on Program Outcomes

Sources of information for identifying outcomes include the sources for identifying a program's mission and objectives (presented under Step 1).

- Legislation and regulations,
- Mission statements contained in budget documents,
- Strategic plans (Department, POC, or program),
- Program descriptions and annual reports,
- Discussions with upper level officials and their staffs,
- Discussions with legislators and their staffs,
- Discussions or meetings with customers and service providers,
- Input from program personnel,
- Complaint information (What have customers complained about?),
- Mission statements used by levels of other government for similar programs.

Additional information on program outcomes can be obtained through a range of activities that can be conducted in a short amount of time.

- Customer focus groups,
- Focus groups of both program staff and local project staff (especially field personnel),
- Meetings and related input from state and local personnel,
- Use of outcome sequence charts or "logic models",
- Role-playing by program staff (acting as customers)

Focus Groups

Focus groups are an excellent way to obtain input from a program's customers as to what quality characteristics and outcomes are important to them. For the Department, customers often will be SEAs or LEAs. For some programs, they might be teachers, parents, or students. Focus groups of program or project personnel, especially personnel who frequently work in the field with customers, can also be used to identify outcomes likely to be of concern to customers.

Exhibit 9 identifies typical steps for focus groups. The participants will likely make many gripes about the program and participants' problems with it. These should be presented to the program manager and staff (anonymously) for possible action. The subject matter of the gripes (e.g., delays in providing program services) and the other comments of the participants concerning what they liked and did not like about the program should be examined to identify program characteristics that represent outcomes to be tracked. The participants' comments and concerns should help the program identify objectives and specific program characteristics for which outcome indicators should be developed.

Exhibit 9
Introduction to a Focus Group

1. Plan the sessions. Determine the information needed, the categories of participants, the timing, location, and other administrative details.

2. Invite approximately 8-12 customers to participate in each focus group meeting.

   These persons can be chosen from lists of customers. The information obtained from focus group participants will *not* provide statistical data. Statistical sampling is not needed. The main criteria is that the participants have had experience with the program.

3. Schedule the meeting for a maximum of two hours. Hold it in a pleasant, attractive, comfortable location. Soft drinks and snacks might be provided.

4. Select a facilitator to facilitate the meeting, one who is experienced in conducting focus groups.

5. After introductions and an overview of the purpose of the meeting, the facilitator should ask the participants the following two questions:
   - What do you like about the service?
   - What don't you like about the service?

   The facilitator can ask these questions in many different ways and should solicit from each participant his or her views on these questions. The facilitator's main job is to establish an open, non-threatening environment and to obtain input from each participant. Facilitators should never debate or argue with the participants.

6. Assign a recorder to take notes on what the participants said.

7. Have the reporter and facilitator summarize the findings from the meeting in writing. The report should extract from the participants' contributions those program outcome-related characteristics that the program should consider tracking.

A variation of the *customer* focus group is a *program or project personnel* focus group. Most of the procedures are the same. However, under item #5, the facilitator would ask:

- What do you believe your customers like about the service?
- What do you believe your customers don't like about the service?

The program or project personnel selected to participate should include a broad representation of personnel likely to be familiar with customer concerns. Personnel who frequently work in the field and first-line staff usually should be included as participants.

BEST COPY AVAILABLE

Focus groups are not costly but require effort to arrange and administer. Participants usually do not need to be paid. Preferably such meetings would be held in a number of locations throughout the country. It is likely to be sufficient, and at least better than doing none at all, to hold them in convenient Washington D.C. area locations, to save funds. The program should use a trained facilitator, if funds are available.

A variation that can be used along with, or if resources are *very* tight instead of, customer focus groups is to use program personnel as the participants (assuming that the program has numerous personnel who are not members of the working group that is developing the outcome measurement process).

Meetings and Related Input From State and Local Personnel or Other "Partners"

Most, if not all, Department of Education programs involve participation by other agencies at the state and local levels. These are usually public or private nonprofit agencies, but also can be the business community (such as with school-to-work programs). External input into the program's outcome measurement process, both the identification of outcomes to be measured and the data collection procedures (especially if some of the data collection involves them) is usually advisable.

The program should attempt to obtain their input through meetings, telephone and mail, conference calls, FAX, Internet, and any other forms of communication.

Caution: If the program seeks input from these other organizations, as is usually highly preferable, the program should take their suggestions seriously and not merely use these communications as a method for being able to say that the program had sought outside input. If you ask for advice, be prepared to use it or face the good possibility that spurned advice will offend those organizations.

Some programs may believe it is preferable to work with other organizations as partners in designing and implementing the outcome measurement process. Such organizations could include state education agencies, school districts, private non-profit organizations (such as parent organizations), and business organizations.

These are situations in which the program believes that desired outcomes would be best achieved by obtaining voluntary agreement among organizations as to: (a) the outcome indicators to be collected; (b) how they should be collected; (c) the short and long term targets for each outcome indicator; and (d) the roles and responsibilities of each organization in providing the particular educational service. The involvement of these organizations from the outset can facilitate future data collection efforts and possibly reduce the costs to the Department of data collection..

Such agreements have been labeled "performance partnerships." This is a new concept, and likely would require considerably time and effort by the Department of Education program to work out with other organizations. For programs, such as school-to-work programs, that are closely related to programs in other federal departments, however, such partnerships may be considerably easier to work out.

Example

For the Eisenhower Professional Development Program, the Department formed a "calling circle" of interested state coordinators. These State coordinators spent a day in Washington discussing the draft indicators that the Department had prepared. They reviewed revisions and discussed draft data collection tools by phone, e-mail, and FAX. The Department also sent the Eisenhower draft indicator process description to all State coordinators for comment and subsequently discussed the indicator system and draft data collection instruments at sessions during the annual national meeting. The "calling circle," with its greater involvement, provided the most in-depth input from States.

Outcome-Sequence Charts/Logic Models

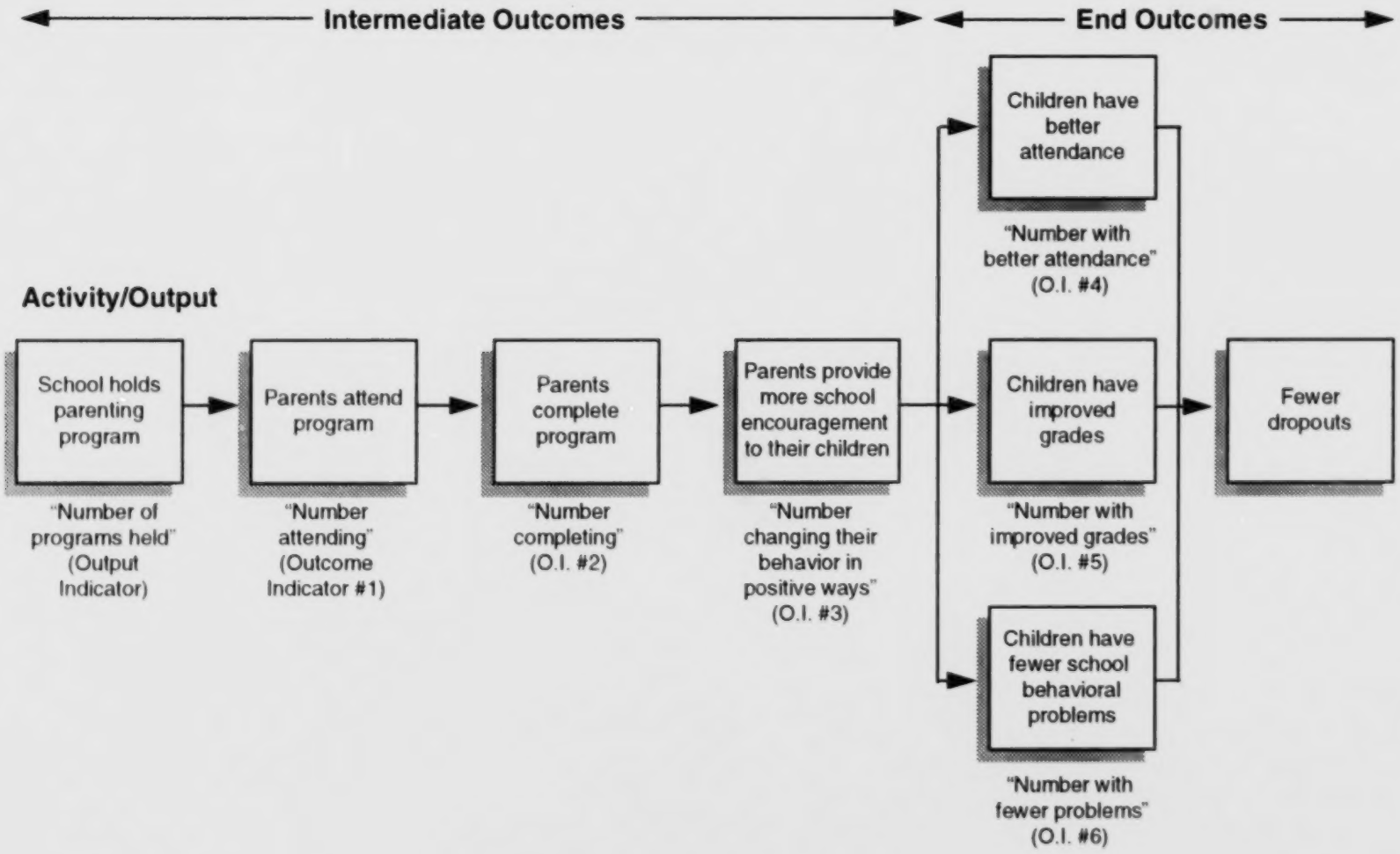
Outcome sequence charts (sometimes called "logic models") can be used to help identify and sort out the various performance indicator categories. A program is likely to find preparation of such charts to be quite helpful. Outcome sequence charts provide a visual depiction of what a program is expected to produce. The charts can trace the anticipated "cause and effects" path of program activities.

Outcome sequence charts should identify the key events that are expected to occur, beginning with program activities, and moving to expected outputs, to intermediate outcomes, and finally to end outcomes. They usually consist of a series of boxes representing program activities, outputs, intermediate outcomes, and end outcomes, that are connected by arrows. These diagrams are another way of identifying and organizing the outputs and outcomes for a particular program.

Exhibit 10 illustrates the sequence of expected events from *program activity* to *outputs* to *intermediate outcomes* to *end outcomes* — for a drop-out prevention/parental involvement program. The program's activity (workload) here is providing programs (classes) to help parents to be more supportive of their children's learning efforts. The exhibit also illustrates specific outcome indicators that might be used to measure each outcome.

The "programs/classes held" are "outputs." The "parents attending these classes" and "completing the program" can be labeled as: "intermediate outcomes." These outcomes will be important *to program managers. They indicate the program's success in attracting* parents and retaining them through the end of the program. The number of parents who attended those program activities and then encouraged their children to learn, indicate that the program actually affected those parents, bringing change that is expected to lead to improved student learning. These changes in parent actions, however, do not tell what outcomes *resulted.* Increased attendance, improved grades, and fewer school behavior problems by the students are the desired "end outcomes." Fewer school dropouts are also hoped for, but this cannot be completely observed, perhaps, for many years. The outcome indicator sequence chart could go even further to include work and earnings histories of the students as longer term end outcomes. Each outcome (and output) on the exhibit should be important to the program and, if possible, included in the program's outcome measurement process.

Exhibit 10. Outcome Sequence Chart Illustrating Activities and Outcomes for Parental Involvement Program


Intermediate Outcomes
End Outcomes
Activity/Output
School holds parenting program
"Number of programs held" (Output Indicator)
Parents attend program
"Number attending" (Outcome Indicator #1)
Parents complete program
"Number completing" (O.I. #2)
Parents provide more school encouragement to their children
"Number changing their behavior in positive ways" (O.I. #3)
Children have better attendance
"Number with better attendance" (O.I. #4)
Children have improved grades
"Number with improved grades" (O.I. #5)
Children have fewer school behavioral problems
"Number with fewer problems" (O.I. #6)
Fewer dropouts

Additional sample outcome sequence charts are shown in Exhibits 11 (relating to systemic reform) and 12 (for the Star Schools distance learning program). The earlier blocks on these charts (those furthest to the left in the exhibits) show program activities and outputs. These usually represent products over which the program and its grantees have fullest control. Those products further to the right are the closest to being end results. The sequencing of the blocks and their products also represents the order in which the products usually occur.

Additional sample outcome sequence charts are shown in Exhibits 11 (relating to systemic reform) and 12 (for the Star Schools distance learning program). The earlier blocks on these charts (those furthest to the left in the exhibits) show program activities and outputs. These usually represent products over which the program and its grantees have fullest control. Those products further to the right are the closest to being end results. The sequencing of the blocks and their products also represents the order in which the products usually occur.

For example, in the Star Schools Program example in Exhibit 12, the program considers that schools making distance learning programming available is an initial outcome. Getting teachers to use distance learning in their classes, and having students participate in classes where distance learning technology is used, represent important and more advanced accomplishments. However, all of these are intermediate outcomes that do not indicate whether improved learning or achievement occurred. Improved learning is the end outcome intended for the distance learning program. Some personnel may feel that the outcome "students report increased interest in school" is an intermediate rather than an end outcome, and it could be so classified. People can legitimately disagree over the category of an item. In situations such as this, the category into which it falls usually does not affect the measurement process.

The members of the program's outcome measurement working group should first individually, and then collectively, construct outcome sequence charts for their program. *The various intermediate and end outcomes identified by the group become candidates for regular outcome measurement.*

Role-Playing By Program Staff

An easy, and even fun, procedure for identifying program outcomes is to have program staff role- play as customers. Individual program staff would each take the role of one of the program's customers—such as representatives of SEAs and LEAs (district and/or school officials), teachers, students, parents, and/or the general public (or whoever have been identified as customers for the program).

This procedure is likely to be particularly valuable if the program is not able to hold customer focus groups. Role playing has the advantage that it helps sensitize program staff (and working group members) to customer concerns—which should help the working group identify customer-oriented outcomes.

Over perhaps an hour, each participant would express their concerns (in their respective roles) on the program. The participants can be asked the same questions posed to focus groups: "What do you like about the program? What don't you like about it?" Each participant would draw on their own knowledge of the program and what their experiences have indicated would likely be the reactions of the customers.

Exhibit 11
Systemic Reform: Illustrative Conceptual Outcome Sequence Chart

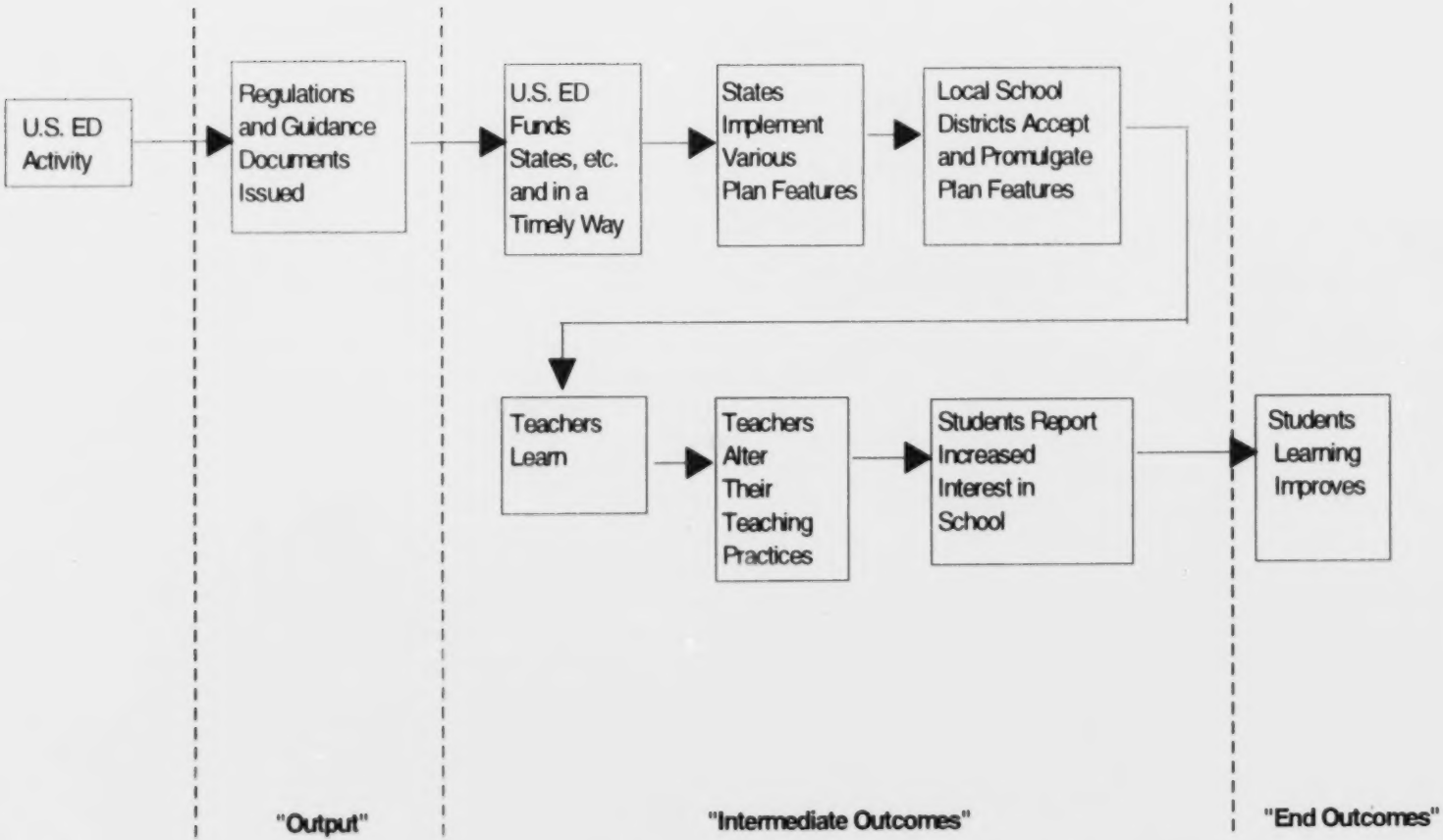
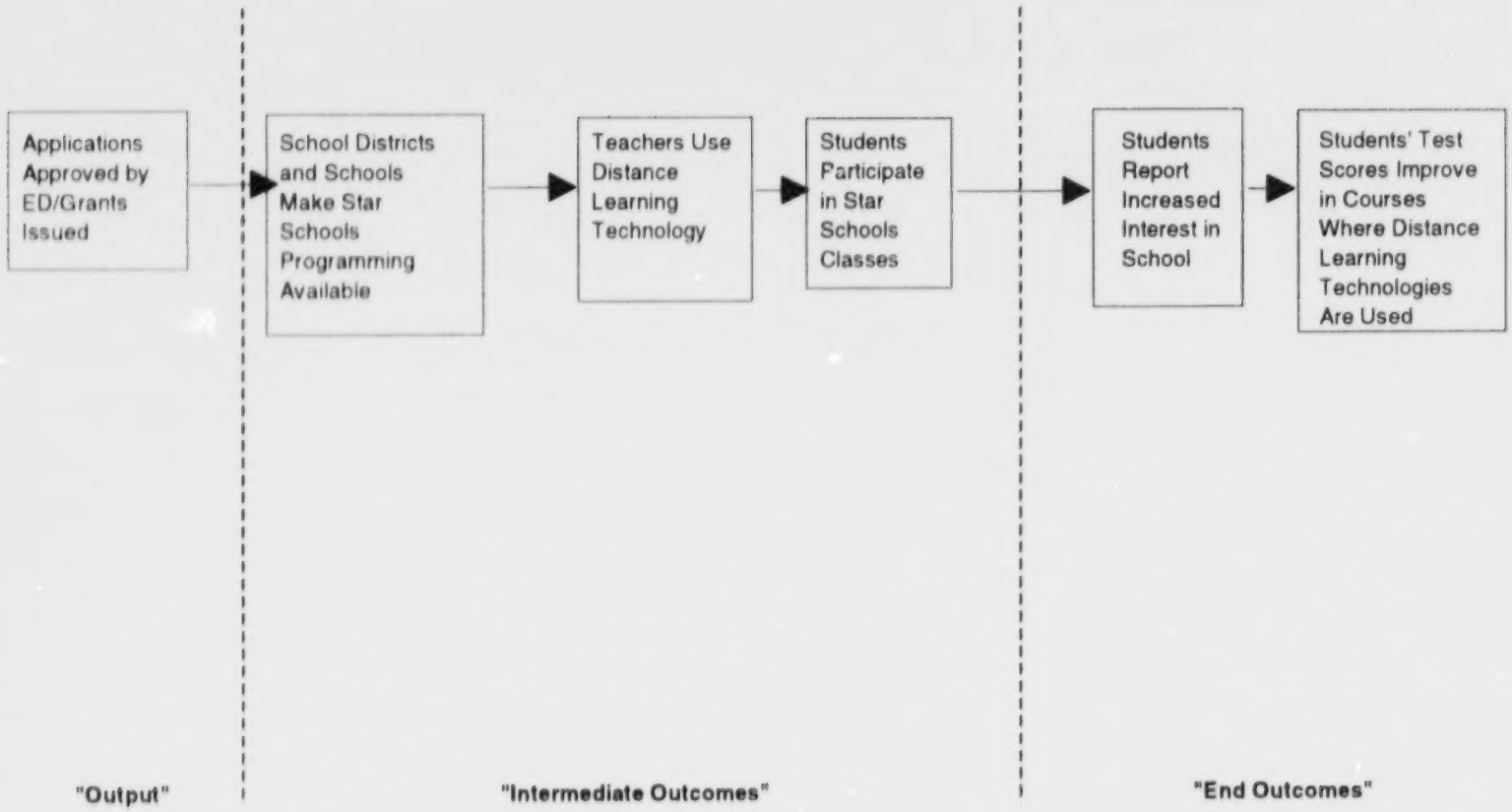

U.S. ED Activity
Regulations and Guidance Documents Issued
U.S. ED Funds States, etc. and in a Timely Way
States Implement Various Plan Features
Local School Districts Accept and Promulgate Plan Features
Teachers Learn
Teachers Alter Their Teaching Practices
Students Report Increased Interest in School
Students Learning Improves
"Output"
"Intermediate Outcomes"
"End Outcomes"

Exhibit 12
Star Schools: Sample Outcome Sequence Chart


Applications Approved by ED/Grants Issued
School Districts and Schools Make Star Schools Programming Available
Teachers Use Distance Learning Technology
Students Participate in Star Schools Classes
Students Report Increased Interest in School
Students' Test Scores Improve in Courses Where Distance Learning Technologies Are Used
"Output"
"Intermediate Outcomes"
"End Outcomes"

Someone should be asked to record the findings of the role-playing session as to the potential outcome characteristics identified during the session. As with focus groups, the note takers would identify the outcomes explicitly or implicitly identified by the role players that should be considered candidates for regular outcome measurement —both intermediate (including indicators of service quality) and end outcomes.

Gathering Candidate Outcomes Identified From All Sources

Before you finish identifying outcomes for your program from the above sources, consider these questions:

1. Do the outcomes cover each element you identified in the mission/objective statement?

2. What bad things would happen to customers if the program's budget and resources were substantially cut or deleted? What would be the consequences? What benefits would customers receive if the program's budget and resources were increased? Think through the implications. This may help to identify other outcomes that have not yet been included in the list of outcomes.

3. Are there any potentially *bad* consequences or effects that are associated with the program and should also be monitored? If these can be tracked on a regular basis, they should be included as outcomes. Some of these may already have been identified in your mission/objective statement as something to be minimized by the program.

4. Put yourself in the role of each category of customer that you identified earlier. Identify the concerns that each category of customer is likely to have regarding the program and its services. What would customers consider to be good or bad service, and why? Include those characteristics in your list of outcomes. Identify whether each characteristic is an intermediate or end outcome.

Gather together into one list all the outcomes that you have identified from all sources. Work out overlaps and duplications. Identify which are intermediate outcomes and which are end outcomes. By this stage, you have probably identified a long list of outcomes. Even though the list may seem lengthy, you probably should not attempt to prioritize or screen out outcomes, except for those that appear truly trivial.

Step 3. Select Outcome Indicators

"Outcomes" are *not* the same as "outcome indicators."

Each outcome that you identify needs to be translated into one or more outcome indicators that identify specifically *what is to be measured.* The specific indicators that will be used will depend in part on the particular data source and data collection procedure to be used. (For example, if service timeliness is being assessed by surveys of customers, the indicator will probably be the "percent of customers giving particular ratings to service timeliness." However, if program records are used to assess timeliness, the indicator will likely be something like the "percent of service requests that exceeded the program's standard for responding.")

An outcome indicator usually identifies a specific *numerical value* which will indicate progress toward achieving an outcome, such as a number or percentage (or ratio).

Note: In this manual, we discuss separately the selection of the target value that the Department seeks to achieve for each indicator. (For a discussion of setting targets for each indicator see the section "Comparisons to Pre-Selected Targets" under Step 6, "Comparing Findings to Benchmarks.") The Department, in such documents as its Strategic Plan, combines the targeted values with the indicators to produce its performance indicators. This is a difference in presentation format, not in the substance of what a program needs to do to produce useful performance information.

Exhibit 13 presents some criteria for selecting outcome indicators. You might rate each indicator on the following criteria. Exhibit 14 provides a checklist you may wish to apply to each outcome indicator.

Exhibit 13.
Some Criteria for Selecting Outcome Indicators

- *Relevance* to the mission/objectives of the program and to the outcome which it is supposed to help measure.
- *Importance* of what it measures.
- The extent to which it might be *duplicated by, or overlap with, other indicators.*
- *Understandability* of the indicator.
- The extent to which the program has *influence/control* over the values of the outcome. But do not overuse this criterion. Often a program will have less influence over the most important outcomes, especially end outcomes. As long as the program is expected ultimately to have some tangible, measurable effect on the outcome, the outcome indicator should be a candidate for inclusion—whether the effects are indirect or direct.
- *Feasibility and cost* of collecting the indicator. However, note that sometimes more costly indicators are the most important—and should be retained. (Data collection procedures and their costs are discussed under Step 4.)

Exhibit 14
A Checklist for Outcome Indicators

1. Does each indicator measure some important aspect of the outcome?

2. Does each indicator included start off with a numerical designation such as "number," "incidence," "percentage," "rate," or "proportion" of . . .?

3. Does your list of indicators cover all the outcomes?

4. Does your list of indicators cover all the "quality" characteristics of concern to customers, such as service timeliness?

5. Does your list of indicators include relevant feedback from customers of the program—relating to the outcomes?

6. Is the wording of each indicator sufficiently specific? Often, some words in the indicator will need to be defined more specifically, perhaps later, with the help of "experts." For example, a program that wants to increase the "number of teachers that have received significant professional development opportunities during the year" will need to define specifically what is meant by "significant" in order to be able to measure the outcome indicator in a meaningful way.

Note: The final choice of outcome indicators for an outcome depends on the data source/data collection procedure. Data sources are discussed in the next chapter.

A full list of outcome indicators recently proposed for the Star Schools Distance Learning Program is presented in Appendix 3. Exhibit 15 is an example of outcome indicators that might be used as a starting set for a Parent Resource Center Program.

Exhibit 15
Illustrative Performance Indicators:
Parent Resource Centers

1. Number of activities / publications / on-site visits / coordination activities, etc. actually produced by the Resource Center.

2. Number and types of parents provided services by each project or by activity within the project, such as referrals to service providers, nutrition advice, basic parenting skills training, counseling, literacy training, etc.

3. Number and percent of parents reporting knowledge or awareness of local parental resource center activities (e.g., Do parents know where and how to access information?).

4. Number and percent of parents reporting satisfaction with each of the following characteristics of the services provided by parental resource centers: a) relevance or appropriateness of assistance; b) timeliness of assistance; c) knowledgeability of staff; d) overall helpfulness.

5. Number and percent of parents reporting that parental resource centers led to their taking a more active role in their child's development or education.

6. Number and percent of parents reporting that they had substantially increased their activity after receiving assistance from Center, on each of the following:
   a. Talked with child regularly [define] about school activities.
   b. Checked homework.
   c. Monitored TV viewing.
   d. Monitored going out with friends.
   e. Visited child's school.
   f. Spoke with teacher or counselor.
   g. Read to child at least 3 times per week.
   h. Other activities, especially pre-K related.

7. Number and percent of students of assisted parents whose academic performance improved after reported (parental) behavior change (based on feedback from parents, teachers, and students.)

Note: Item #1 is an output indicator. Items #2 through #6 are intermediate outcomes. Item #7 is an end outcome.

Use of Outcome Sequence Charts

Under Step 2 we discussed the usefulness of outcome sequence charts and provided examples of sample logic models for systemic reform and the Star Schools Program. Exhibits 16 and 17, at the end of this chapter, recreate these models but include selected performance indicators. Exhibit 10 presented in Step 2 also includes illustrative indicators for a drop-out prevention/parental involvement program.

Tips on Selecting Outcome Indicators

> **Tip 1:**
> **You are likely to identify a large number of possible outcome indicators. At this stage, avoid attempting to reduce the number.**

This list of indicators is a list of *candidates*. Avoid discarding indicators at this stage because of the belief that the data collection would be infeasible or too expensive. Later, when other personnel review the list, you may collectively decide that the list needs to be reduced and choose which indicators are of less importance. Some indicators may be needed primarily for *internal* program tracking purposes. A smaller number of indicators might be extracted for reporting performance *outside* the program. (A more detailed discussion of developing indicators for internal versus external reporting purposes is presented in Step 8.)

Exhibit 18 is an extract of 12 key indicators from the complete list of 34 outcome indicators (in Appendix 3) that were proposed for the Star Schools distance learning program. These 12 might be the ones used for external reporting (to the Department and OMB). The remaining 22 would primarily be used for internal program management.

> **Tip 2:**
> **Don't exclude an outcome indicator merely because the program has been doing very well for a long period of time. Take credit for this. Include *any* outcome indicator that is an important outcome for the program.**

For example, suppose a national survey of parents (administered in three consecutive years) consistently shows that greater than 95% of those surveyed report awareness of the Department's Family Involvement Partnership for Learning. This consistent positive response is not a sufficient reason for excluding such an indicator.

Take credit for good accomplishments!

Options for Statistical Forms of the Indicators

Outcome indicators are often expressed as the number (incidence) or percentage (proportion or rate) of something. Often, a program may want to include *both* forms, for example, the number and percent of children that passed a criterion referenced test.

Percentages can be expressed in a number of ways, including:

1. The percent that fell into one particular outcome category, such as the percent that rated some service characteristic as "good."

2. The percent that fell above (or below) some targeted value.

3. The percent that fell into particular outcome intervals, such as the percent that fell between the 50th and 75th percentiles.

Some indicators are expressed as something that the program wants to *maximize*. Others are expressed as something the program wants to *minimize*. Often, you have a choice of which form you want. Do you want to report the glass as half full or half empty? For example, you could choose the percentage of students that reported using illegal drugs during the past month, or alternatively, the percentage that reported not using illegal drugs. The first form is expressed as something the program seeks to minimize; the second form as something to be maximized.

Another example: An outcome indicator measuring "customer satisfaction" might be either (a) percentage of respondents that rated a particular service characteristic as either excellent or "good" (to be maximized) or (b) percentage of respondents that rated the particular outcome characteristic as either fair or "poor" (to be minimized).

A Useful Option: Identify the Extent of Program's Influence/Control Over Each Outcome Indicator

For most (if not all) outcomes, the program is likely to have only partial control over the value of the indicator. In almost all cases, external factors beyond the control of the program will also affect indicator values. This is particularly so for *end outcomes*, but will likely also apply to *intermediate outcomes*.

As long as the program can have some effect on the value of an outcome indicator, the indicator should be a candidate for inclusion.

You might want to identify and report the approximate degree of control the program has over each outcome indicator. This will alert users of performance reports about the extent to which the program can affect each indicator.

If so, assign approximate values to the program's degree of control over the outcome indicators. You could, for example, assign categories to each indicator as to the program's degree of control, such as "little," "some," or "considerable" control. If so, however, each category should be defined as specifically as possible so that others can properly interpret these indicator categories.

Exhibit 16
Systemic Reform: Conceptual Outcome Sequence Chart
Using Illustrative Performance Indicators

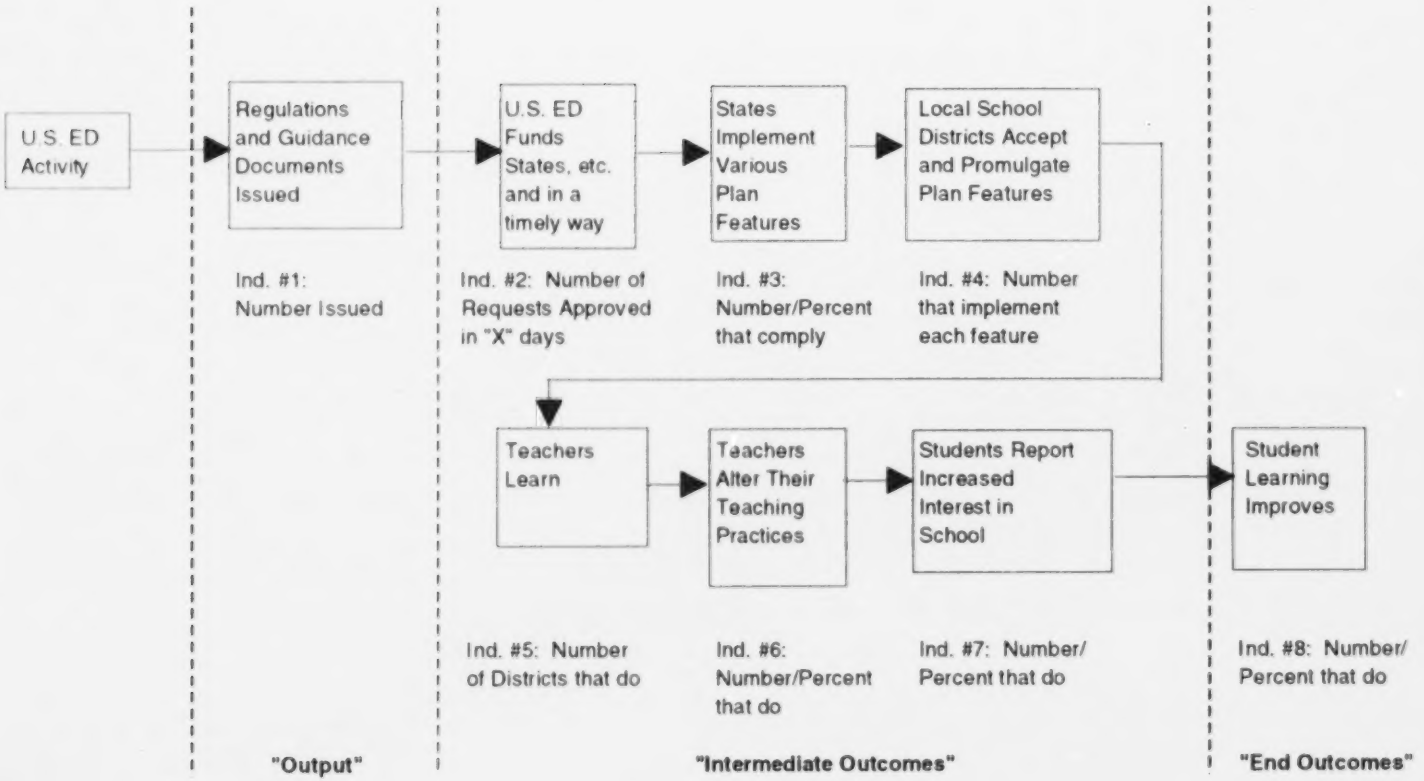

U.S. ED Activity
Regulations and Guidance Documents Issued
U.S. ED Funds States, etc. and in a timely way
States Implement Various Plan Features
Local School Districts Accept and Promulgate Plan Features
Ind. #1: Number Issued
Ind. #2: Number of Requests Approved in "X" days
Ind. #3: Number/Percent that comply
Ind. #4: Number that implement each feature
Teachers Learn
Teachers Alter Their Teaching Practices
Students Report Increased Interest in School
Student Learning Improves
Ind. #5: Number of Districts that do
Ind. #6: Number/Percent that do
Ind. #7: Number/ Percent that do
Ind. #8: Number/ Percent that do
"Output"
"Intermediate Outcomes"
"End Outcomes"

Exhibit 17
Star Schools: Outcome Sequence Chart
Using Illustrative Performance Indicators

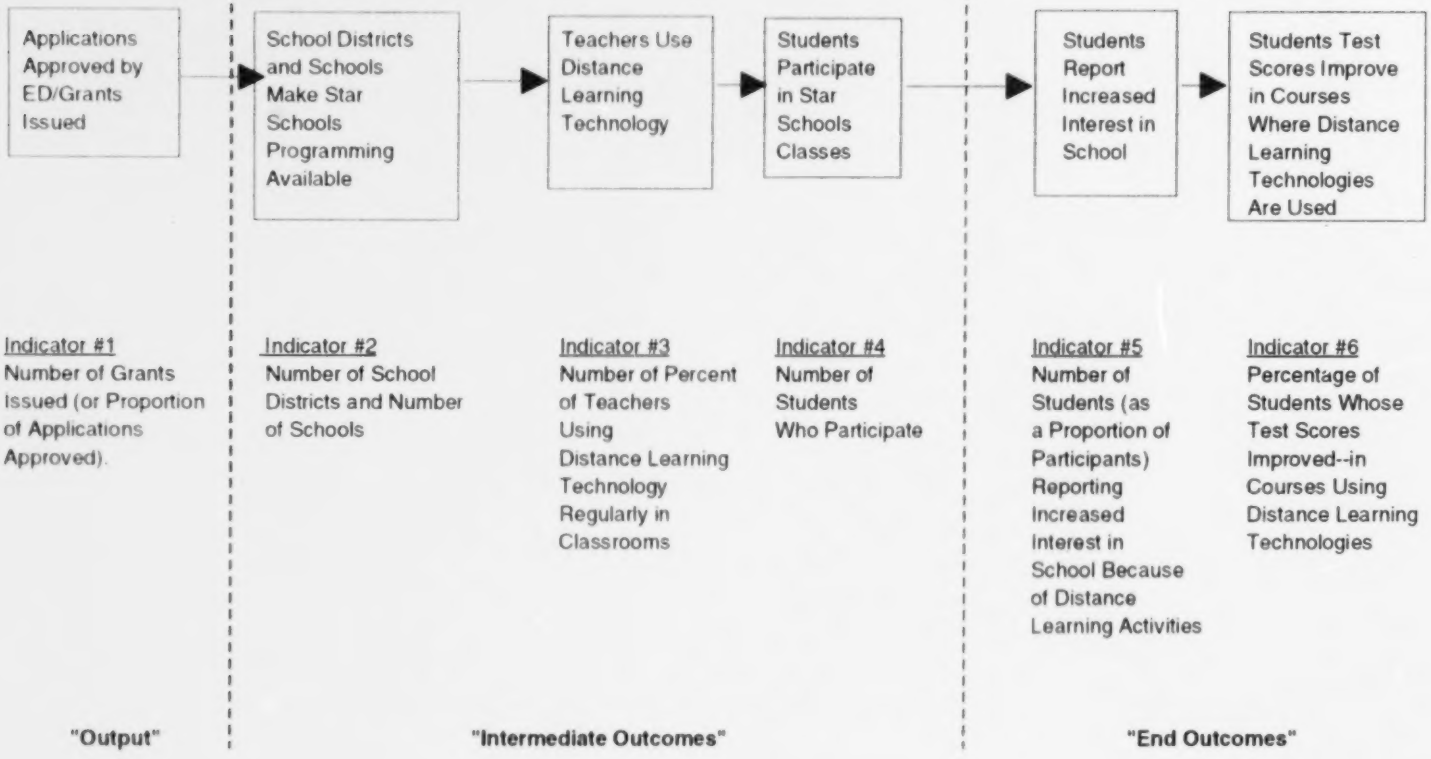

Applications Approved by ED/Grants Issued
School Districts and Schools Make Star Schools Programming Available
Teachers Use Distance Learning Technology
Students Participate in Star Schools Classes
Students Report Increased Interest in School
Students Test Scores Improve in Courses Where Distance Learning Technologies Are Used
Indicator #1
Number of Grants Issued (or Proportion of Applications Approved).
Indicator #2
Number of School Districts and Number of Schools
Indicator #3
Number of Percent of Teachers Using Distance Learning Technology Regularly in Classrooms
Indicator #4
Number of Students Who Participate
Indicator #5
Number of Students (as a Proportion of Participants) Reporting Increased Interest in School Because of Distance Learning Activities
Indicator #6
Percentage of Students Whose Test Scores Improved--in Courses Using Distance Learning Technologies
"Output"
"Intermediate Outcomes"
"End Outcomes"

Exhibit 18
Selected Outcome Indicators for External Reporting Purposes:
Star Schools Program

Intermediate Outcomes[a]

1. Number and percent of (a) K-12 students (ethnicity, age/grade level, gender); (b) teachers (ethnicity, level of experience); and c) others participating in the program. [#4]

2. Number and percentage of K-12 students who are disadvantaged, LEP, or have disabilities, who are participating in the program. [#6]

3. Number of students enrolled in Star Schools high school credit courses, college preparatory courses, or advanced placemet courses that had not been available previously. [#11]

4. Number of school administrators reporting changes/improvements in classroom teaching and learning practices, related to distance learning. [#13]

5. Number and percent of teachers reporting that the learning materials for students were an effective means to adequately cover the subject topics. [#15]

6. Number and percentage of schools reporting that the distance learning materials they had received were (a) understandable, (b) adequate/complete, and c) being used. [#20]

7. Number of schools that continue to use distance learning services (courses, staff development) when project is no longer receiving Star Schools funding, by characteristic of activities that continue. [#25]

End Outcomes[a]

1. Percentage of students who report that the distance learning activities added significantly to the quality of the course or subject, by various student characteristics, such as age, gender, ethnicity. [#28]

2. Percentage of students reporting increased interest in school because of distance learning activities in their classes. [#30]

3. Percentage of teachers reporting: (a) increased learning by their students; (b) improved student attendance; c) increased interest in subject area; (d) improved critical thinking and problem solving, attributable at least in part to the use of distance learning activities in their classes. [#32]

4. Percentage of students whose test scores improved significantly in courses in which distance learning technologies had been introduced and were a significant part of the instruction. [#33]

5. Number of non-college bound students who are employed within (a) six months or (b) one year of completing high school and who report that participation in the program was a contributing factor in being selected for employment. [#34]

a. The numbers in brackets represent the indicator number used in the full set of proposed outcome indicators for the Star Schools Program found in Appendix 3.

Step 4. Identify Data Sources and Data Collection Procedures

Major Sources of Outcome Data

A major step is to identify sources of data for each indicator. Until you have identified a reasonably practical way to collect the data, do not consider your list of indicators and breakout information (discussed in Step 5) to be firm.

The four major sources of education outcome data are:

1. Program and agency records.
2. Administered tests (usually of students).
3. Customer surveys.
4. Trained observer ratings.

These sources and recommendations for their use are each discussed below, followed by a section on choosing among these data collection procedures.

Outcome Information Obtained from Program and Agency Records

Depending on the program, data might be obtained from SEA records, LEA records, project records, and so on. Here are some examples of outcome information that might be obtained from such records:

- Data on timeliness/response times such as timeliness of responses to requests for waivers, or information on student loan repayments (intermediate outcomes).

- Number of voluntary users of a particular service who completed the service or program offered, such as number of parents who completed a program to help them help their children with their education (intermediate outcomes). If no such records are kept, usage of services can be obtained from other sources such as surveys.

- Number of complaints received—preferably broken out by the subject of the complaint (intermediate outcomes). Programs may have to add procedures to record the data. (A program might prefer to track the "number of *valid* complaints." This would require the program to define "valid" and to have staff determine in a reliable way whether each complaint was valid.)

- Number of states or school districts that have implemented various aspects of educational reform, such as new curricula and performance standards aligned with new assessments (intermediate outcomes).

- Dropout rates (end outcome).

- Absenteeism rates (intermediate/end outcome).

- Incidence of student disturbances within the schools (end outcome).

- Incidence and rates of identified teenage illnesses and deaths (end outcome).

- Results of test scores (end outcome).

- Various demographic characteristics of the students (breakout/explanatory characteristics—see Step 5).

Note that in addition to being a source of outcome information, agency records are also the main source of data on the amounts of *input*, both dollars and employee time, and amounts of *output* produced by the program.

Advantages of Agency Records

- Attractive because of availability and low cost.
- Procedures are usually familiar to program personnel.

Disadvantages of Agency Records

- Modifications to existing record-collection processes will often be needed to obtain outcome data. For example, though collecting response time data is common for some services, other programs are likely to have to modify their procedures to generate these data. A program will have to:
  - Record the time of receipt of a request for service.
  - Define when "completion" of the response has occurred.
  - Record the time of completion of the response.
  - Establish data processing procedures to calculate and record the time between these two events.
  - Establish data processing procedures for combining (aggregating) the data on individual requests—to provide the data needed for outcome indicators.
- Records alone seldom provide enough information on program quality and outcomes.
- Outcome information will sometimes have to be obtained from records of other programs or other agencies (such as SEAs and LEAs).

Administered Tests

Assessments of student performance through various forms of tests are a major source of outcome data for many programs. (As noted above, they can also be considered a type of agency record.) For many education programs test results are considered major outcome indicators. For outcome measurement purposes, grouped test data categorized by various breakout characteristics (see Step 5) can be major end outcomes for many Department programs.

It is beyond the scope of this manual to cover the major pros and cons of the many forms of assessment/test procedures currently available. The key data collection issues for programs are the following:

- The availability of appropriate test results, especially for the particular clients of the program—testing by schools may occur only for some grades and cover only some subjects;
- The accessibility of that data from individual schools; and
- The added cost if special testing is needed to cover special material or more frequent data—testing can be quite expensive.

While test scores can be a very valuable source of end outcome information, obtaining test results for regular and timely outcome measurement can pose considerable difficulties to a program. If such difficulties are insurmountable, the program will likely need to rely on surrogate indicators such as program customer surveys (discussed below) to obtain their perceptions of learning progress or on intermediate outcome indicators, which are usually more readily available.

Customer Surveys

Customers are an important source of information about program outcomes. Students and their parents are ultimately the primary customers for U.S. Department of Education services. Also, SEAs and LEAs are recipients of services, as are teachers and parents, and thus are often customers of Department services. Customer surveys will usually be surveys of only those agencies or persons to whom the program has provided services directly or indirectly. In some special cases a program may want to survey all agencies or persons whether or not served.

Surveys of customers, systematically conducted, are a major way to obtain information on outcomes such as customer behavior and satisfaction with various program characteristics. Informal ways to obtain customer feedback usually do not provide statistically valid data. Complaint data, while useful, do not cover the full range of information on service performance. Also, complainers probably are not representative of the full population of those served. Focus groups, though a very good means of helping to identify what outcomes should be measured and to interpret data findings, do not provide reliable statistical data.

Exhibit 19 lists the various types of information that programs can obtain from customer surveys.

Exhibit 19.
Information Obtainable from Customer Surveys

- Ratings of overall satisfaction with a service and of the results achieved.
- Ratings of specific service quality characteristics.
- Data on actual customer experiences and results of those experiences.
- Data on customer actions/behavior sought by the program's services.
- Extent of service use.
- Extent of awareness of services.
- Reasons for dissatisfaction or non-use of services.
- Demographic information about customers.
- Suggestions for improving the service.

Advantages of Customer Surveys

- Much of the information listed in Exhibit 19 is unavailable from other sources.
- Surveys provide input from the program's customers, adding meaningful information and credibility

Disadvantages of Surveys

- They are often unfamiliar to agency personnel and require special expertise, at least in the initial design of the survey (and, if telephone or in-person interviewing is used, to administer the interviewing process).
- They can be costly.
- The evidence obtained from them is based on respondents' perceptions and memory and may be less convincing than data obtained from other sources such as school records.

Content of Customer Surveys

Questionnaires used for outcome measurement should include:

- **Questions relating to the outcomes of services.** The responses should be used to develop indicators of quality and outcomes relating to their experiences relative to the services provided to them. Seek both ratings and factual information from respondents. Questions usually should be asked about both *specific* service characteristics and *overall* service ratings. (This will help program personnel identify specific service problems.)
- **Questions seeking information about the type and amount of the service that the respondent used**. This information can be used to relate service outcomes to the type and amount of services that respondents received. Also, ask about the extent of awareness of the services.
- **Diagnostic questions**. Ask why respondents gave particular answers or ratings. Particularly, ask them to explain poor ratings. (Provide the responses—anonymously—to program personnel to help them identify needed improvements.)
- **Questions seeking demographic information**. These let the program tabulate (breakout) responses by specific characteristics of customers (such as grade level, age, gender, school lunch participation status, disability status, race/ethnicity, state, urban vs. rural vs. suburban character of school system, etc.)
- **Requests for suggestions for improving the service**. Such a request should usually be included at the end of the questionnaire. (These suggestions should also be provided--anonymously—to program personnel to help them identify desirable improvements.)

Examples of Customer Survey Questionnaires and Their Use

Appendices 1 and 2 provide examples of outcome-oriented questionnaires for the Star Schools Program for administration to teachers and students, respectively. The teacher questionnaire in Appendix 1, for example, asks the teachers to rate various characteristics relating to teachers' experiences with the distance learning materials, such as the adequacy of the equipment and

instructional support materials (Questions 6-10). It also asks respondents to rate the use of distance learning in professional development activities in which respondents participated (Question 13). All these questions provide data for intermediate outcome indicators.

Other questions ask respondents to assess the extent to which the distance learning effort contributed to increased student learning and interest in the course subject matter (Questions 11 and 12). These questions provide data for assessing the *end outcomes* of the program.

The teacher questionnaire also provides descriptive information about the teachers and the extent of the teachers' use of distance learning procedures in the past year (Questions 1-5). These questions are for use in comparing the outcomes identified by the teachers to these characteristics, for example, to assess the extent to which increased student interest and learning were related to the amount of distance learning provided by the teacher.

Both questionnaires in the appendices also ask respondents for explanations for poor ratings and suggestions for improving the services. Such responses should be provided to program and project personnel (without identification of respondents)—for their use in improving the service.

Customer surveys have typically been used to assess customer satisfaction but also can, and should, be used to help assess extent of improvement.

For example, question 10 in the teacher questionnaire in Appendix 1 focuses on *overall satisfaction*. As worded ("Overall, how would you rate the helpfulness of [distance learning program] to you in your teaching?"), the question does not ask about the results of the service (thus, it provides information on intermediate outcomes). The questionnaire also asks for information about *student improvement*, which is an end-outcome indicator (Questions 11 and 12). In fact, these questions address the issue of *causality* from the customer's perspective by asking for the teachers' assessments of the extent to which favorable outcomes were "because of" the program service, in this case, distance learning programs. (The question asked in #11 is "To what extent do you believe your students were able to improve their learning because of the use of your distance learning program: not at all, a little, somewhat, or considerably?")

A slightly different approach is to ask first about the extent to which the outcomes were favorable and then to ask about the extent to which the program contributed to the favorable outcomes. The responses from both questions can then be combined to provide data for an end outcome indicator such as "percentage of clients reporting particular degrees of improvement *and who also* reported that the program contributed significantly to the changes reported."

These questions about the extent of improvement and the extent to which the program caused that improvement provide *respondents' opinions*. Such evidence of improvement and program impact will usually be less convincing to users of outcome reports than "hard" evidence from, say, agency records and test scores. Nevertheless, such information is relevant and, if test score data cannot be obtained, may need to be used to provide the best available information on learning improvement.

Customer Survey Administration Methods

The major ways to administer surveys are:

- **Mail**—inexpensive, but requires special procedures to obtain acceptable response rates.

BEST COPY AVAILABLE

- **Telephone**—a good process but requires considerable interviewer time (and interviewer training).
- **In-person** at the person's home—expensive and not likely to be feasible for regular (e.g., annual) data collection.
- **Administration at a public facility**—inexpensive, but in some cases may not be fully adequate for assessing outcomes. For example, ratings of teacher-training sessions sponsored by a Federal program obtained at the end of the training sessions at the training sites can provide information on the way the training was conducted and its content. However, such immediate ratings are not likely to be helpful in indicating whether the training actually turned out to be helpful to the teachers in their teaching activities. Later surveys of those teachers would be needed. A useful approach for obtaining post-program survey data, however, is to distribute questionnaires to students or teachers in school buildings—and then have them collected by, or mailed back to, personnel representing the Federal program.
- **Combinations of the above**. This is often likely to be appropriate, using inexpensive mailings supplemented by telephone calls to at least a sample of non-respondents to the mailings.

Key Concerns in Selecting Survey Methods

Three principal issues need attention when choosing the mode of questionnaire administration:

- Response rates
- Accuracy of responses
- Cost

Choosing a survey method involves tradeoffs among these concerns.

At-home, in-person surveys may provide high response rates and detailed information, but are usually too costly.

In-person surveys at a public facility are much less costly and can obtain detailed information, including that about past experiences with the program. This option, however, is *not appropriate* when the program wants information on outcomes that occur *after* the customer leaves the facility—for instance, to assess the usefulness of the program's services after customers leave the program. However, in-school surveys of students and teachers likely to still be at the school at the appropriate follow-up time for assessing a program's outcomes can be a useful option.

Telephone surveys are a less expensive alternative for reaching households than in-person interviews and can achieve good response rates. However, they require considerable interviewer time, and interviewers need training to conduct interviews.

Mail surveys are usually the least expensive. Second and third mailings or telephone reminders to non-respondents will be needed to obtain response rates high enough to provide reliable information. Mailed questionnaires need to be short and uncomplicated. They are not useful for respondents with literacy problems. Mail surveys of SEA and LEA customers by the Department are, however, likely to achieve good response rates, much higher than if a private firm mails questionnaires to random households. Suggestions for increasing response rates to mail surveys are given in Exhibit 20.

Exhibit 20.
Suggestions for Increasing Response Rates to Mailed Customer Surveys

1. The transmittal letter should be signed by a high-level official. For surveys of SEAs and LEAs, the transmittal letter should be personalized and addressed to a specific individual.

2. The transmittal letter should be carefully worded to encourage response. It should emphasize the Department's need for the information from clients in order to improve services in the future. (A sample transmittal letter is presented in Exhibit 21).

3. Mail an advance brief notice (perhaps via a postcard) that the recipients will shortly be sent a questionnaire and asking for their help in completing and returning the questionnaire.

4. The questionnaire should be as short and simple as possible. Questionnaires that are complex, cluttered, or more than four or five pages long should be avoided

5. A stamped, self-addressed return envelope should be enclosed with the questionnaire at each mailing.

6. The transmittal letter should guarantee that responses will *not* be attributed to the respondent or their organization in any report.

7. The questionnaire should be as attractive as possible. Preferably, it should be typeset and printed on good-quality paper. Cutting corners such as using standard office copiers should be avoided.

8. Use two or three mailings. One mailing will not be enough. Telephone and post card reminders also can be used, especially for small samples such as surveys of SEAs.

Exhibit 21.
Sample Transmittal Letter

Dear :

The [program name] is attempting to improve its services to [name the client category]. Your responses to the enclosed questionnaire will provide important information to help us judge the current quality and usefulness of our services and make improvements.

Completing the questionnaire should take about 15 minutes.

Your response will be held in confidence. Results will be reported only in aggregate form. No individual responses will be identified

Please return the questionnaire by [date] in the stamped self-addressed envelope we have provided.

Many thanks in advance for your help. Please call [name and phone number] if you would like more information on this survey.

Sincerely,

High-Level Official

Whatever survey method used, we recommend that you *target at minimum a 50% response rate*. Try to get completed questionnaires from a majority of those persons from whom you are seeking responses. This rate is lower than ideal, but 50% should be adequate for most annual outcome measurement work and can substantially reduce the costs of regular surveys.

Cost of surveys depends considerably on the number of persons being surveyed and the frequency with which they are surveyed, as well as the mode of administration (including the effort made to increase response rates). Some programs may find it feasible to survey *all their customers*, such as by routinely mailing questionnaires to each customer at a specified time after service has been given (and following up with non-respondents at least once). This will apply when the persons to be surveyed are SEA representatives and responses are needed from only a small number of SEA representatives in each state.

When very large numbers of schools, school districts, students, teachers, or parents fall into the population served by the program, sampling the population will usually be needed. The program will then have to decide on the sample size. Larger samples will be needed if the program needs considerable precision, but high levels of precision are seldom likely to be needed. Larger samples will also be needed to the extent that the program wants outcome information on a large number of breakout categories. If, for example, the program wants outcome data on each of the states, the program will need to have large enough samples of respondents in each state to provide the desired information at the desired level of precision.

Cost Saving Ideas

Here are some ways to reduce survey costs:

- Use government personnel where possible and appropriate to oversee the survey work. (Do *not* use persons who are delivering the service as interviewers. This undermines survey credibility.) If sample sizes are small (such as surveys of SEAs), do the surveys in-house.
- Use already available questionnaires. Note that after the questionnaire has been developed it can and (for comparability) should be used to obtain data for future reporting periods. (Inevitably, the program will want to make some changes in questions and question wording from year to year. As long as these changes are not extensive, this should have only minor effects on year-to-year comparability of data.)
- Use inexpensive technical consultants (perhaps from within the Department such as NCES) to help design the survey and the questionnaire.
- Use commercially available software for tabulations.
- Use mail surveys, but do second (and third) mailings.
- Use samples, if necessary, rather than survey everyone.
- Use smaller samples. Avoid excessive precision (99% confidence limits are overkill for Department programs; 95% are well-accepted, but even this level may be excessive for most programs; 90% confidence limits may be fully adequate and reduce the needed sample sizes.)
- If possible, use volunteers to administer surveys. This may be quite practical if the surveys are being administered by nonprofit grantees, who might use community organization partnerships to help with the surveys.

Wording of Customer Survey Questionnaires

It is surprisingly easy to unintentionally include biased or unclear wording. Therefore:

- Always have a professional survey "expert" finalize (or at least review) question wording before the questionnaire is finalized.
- Always pre-test a questionnaire with some customers before full use.
- Mail Questionnaire Format

Here are some suggestions on questionnaire format and style:

- Customers (whether SEAs, LEAs, teachers, students, or parents) are more likely to complete and return questionnaires that look attractive!
- Format and arrange questionnaires so they are clear and easy to handle. Go for a professional, polished look—even if you produce the questionnaire in house.
- Use question response categories that only require respondents to check off which response applies; keep the number of "open-ended" questions to a minimum (questions requiring the respondent to use their own words). However, a small number of open-ended questions (about reasons for responses and suggestions for improving services) can be very informative.
- Consider using colored paper to attract attention.
- Limit the use of skip patterns (in which respondents are given instructions to skip certain questions, depending on their response to a previous screening question); where skip patterns are needed, make them easy to follow.

Tips On Designing and Administering Surveys

Here are some suggestions:

- To develop the content of the questionnaire, establish a working group that includes key service agency representatives, perhaps a representative of the chief administrative officer, a survey expert, and a representative of the program's customers.
- Use an expert in survey development, particularly for selecting samples, questionnaire wording, and training interviewers (if interviewers are to be used).
- For mailed questionnaires, always provide respondents with a stamped, self-addressed return envelope.
- Review and pretest surveys to screen out such problems as:
  - Long, awkward, or ambiguous questions
  - Confusing or incorrect instructions
  - Redundant questions
  - Wording that may offend or sound foolish to respondents
  - Illogical or awkward sequences of questions.
- Translate questionnaires into foreign languages if substantial numbers of limited-English-speaking persons are to be surveyed.
- Over the long run, if the program has the resources, contracting for the regular surveys will make administration much easier for the program than administering the surveys itself. Some suggestions as to what should be included in survey contracts are provided in Exhibit 22. Surveys of small numbers of customers, such as of the states and territories, however, probably can be handled by the program itself, particularly if questionnaires are mailed.

Exhibit 22.
Elements that Should be Included
In a Contract for a Customer Survey

1. The size of the samples, including minimum sizes for each major category of customer for whom data are sought;

2. Survey administration details such as whether the mail, telephone, or both are to be used; the number of mailings or number of follow-up telephone calls, including the minimum completion rate; and the time between mailings and telephone follow-ups;

3. The role of the contractor in developing the questionnaire;

4. The role of the contractor in pretesting the questionnaire (the agency may want to do some of its own pretesting, but the contractor should also do some);

5. Maintenance of confidentiality of responses;

6. Special coding to be done by the contractor (e.g., transforming school district data provided to the contractor into a small number of regions);

7. Specification as to how tabulations are to be handled, such as whether "no answers" and "don't knows" should be included in the denominators for the percentages that are calculated;

8. Products to be provided to the government and in what formats (products should include at a minimum: multiple cross-tabulation tables for each questions, frequency counts for each question, and a fully legible printout of the input data for each returned questionnaire, a detailed description of the survey procedures used and response rates);

9. The time schedule for the work; and

10. Cost.

If the program uses an evaluation contractor (such as the Star Schools Program has done), the contract might include support by the contractor for these surveys.

Last Notes on Customer Surveys

This section has presented only a brief overview of customer surveys. It has not attempted to address in any depth important technical issues such as:

- Sample size
- Sample selection
- Frequency of administration
- Administration procedures
- Obtaining accurate, reliable, data
- Obtaining adequate response rates
- Analysis of results

References for more detailed information on surveys are provided in at the end of the report.

Trained Observer Procedures

Trained observers can be used to rate a variety of outcome conditions that can be seen by the eyes, or other physical senses, of an observer. You can think of this as a variation on building safety, health, and sanitation inspections.

A high degree of accuracy and reliability can be maintained if you have:

- A clearly defined rating system,
- Adequate training and supervision of the observers and the process,
- A procedure for periodically checking the quality of the ratings.

The goal is for different observers, at different times, to give approximately the same ratings to similar conditions.

Applications of Trained Observer Ratings

To apply trained observer ratings to a particular program outcome, the outcome should:

- Have an outcome that can be measured by physical observation,
- Be one that can be rated on a scale that identifies variations in condition.

Examples of applications include the following:

- Ratings of student responses on open-ended test questions and ratings of school projects,
- Condition of facilities such as school buildings,
- Presence and use of special equipment (such as appropriate computers and distance learning equipment),
- Ratings of classroom procedures, such as use of new equipment, by classroom observations,
- Quality of food provided to school children (in this case other senses such as taste and smell can also be used),
- Accessibility of handicapped students to facilities and equipment.

Advantages of Trained Observer Procedures

- They can provide reliable, reasonably accurate ratings of conditions that otherwise are difficult to measure.
- The data can usually also be used to assist the program in allocating its resources throughout the year, if the ratings are done periodically,
- They can usually be presented in an easy-to-understand form to public officials and to the public.

Disadvantages of Trained Observer Procedures

- This is a "labor-intensive" procedure that requires significant amounts of time, including time for training observers.
- Ratings need to be periodically checked to ensure that the observers are adhering to the procedures.
- It is not a common procedure so program personnel may not feel comfortable with it.

Trained observer procedures have limited application to Federal education programs. For more details on these procedures, see Appendix 4.

Identifying Data Collection Procedures

The choice of outcome *indicators* for an outcome depends in part on the data source. Outcome indicators should *not* be considered final until a data source and particular data collection procedure have been chosen. The data collection procedure affects the specific outcome indicator used to measure a particular outcome.

For example, indicators of the intermediate outcome "timeliness" can be obtained from agency records or customer surveys. The indicators obtained from *agency records* might be:

> *The number and/or percentage of requests for which the recorded time between receipt of the request to provision of the service was less than "X" minutes.*

The timeliness indicator using *customer surveys* might be:

> *The percent of surveyed customers who rated the timeliness of the service as "excellent" or "good" rather than "fair" or "poor."*

Consider using more than one procedure and thus more than one indicator to track an outcome. Each procedure, and each specific performance indicator, will likely provide a *different* perspective on that outcome. For example, customer surveys can provide ratings that represent *customer perceptions*. Agency record data and trained observer information provide more *factual* information about the outcome. Both are likely to be relevant for a full perspective on a program's success.

Here is another example. To assess the outcome for school drug abuse prevention, a program might have identified as the major desired outcome: "reduced use of illicit drugs by youth." Agency records and surveys of clients can both be used to measure what reduction, if any, has occurred among participating students. The indicator based on agency records might be the following:

> *The number and percent of youth who had participated in the program who came into either the criminal justice or school systems with a drug-related problem within 12 months after completing the drug abuse prevention program. This information might come from a number of agencies, such as police, courts and hospital records, as well as from school records.*

The youth can also be surveyed. Doing so would yield an outcome indicator such as:

> *The number and percent of students who participated in the program and who reported 12 months after completing the drug abuse prevention program, no drug-use. (Respondents' willingness to admit to drug use is a problem in the use of customer surveys on sensitive topics such as illegal drug use. However, properly conducted surveys have provided what appear to be reasonably honest responses, at least from most respondents.)*

At present, no appropriate *trained observer* procedure seems to be sufficiently useful for detecting drug use.

When making choices among data collection procedures (and the associated outcome indicators) a program will have to trade off cost against precision and accuracy. We suggest that it is better to be roughly right than precisely ignorant.

Step 5. Select Outcome Indicator Breakouts

Transforming the Data Into Really Useful Information

Producing data does not mean that it will be used or be useful. Four elements are needed to transform the data into really useful information to users of that information:

- *Breakouts* of the outcome data for each indicator.
- *Comparisons* of the program's data to other benchmark data.
- *Explanations* as to why the data are the way they are—particularly when the data do not meet desired target levels.
- *Clear presentation of the information* in understandable, useful formats.

This step discusses and suggests what *breakouts* of the outcome data are likely to be useful for your program. Note, however, that while this topic is discussed after data collection procedures, in practice, decisions on appropriate breakouts should be made before finalizing decisions on data collection procedures, so that procedures are selected to provide the needed breakout data.

Comparisons are discussed under Step 6. *Report formats* and *explanations* are discussed under Step 8.

Outcome data should help program personnel identify *where* the program is doing well and where it is not. Including breakouts will make the outcome information *much more useful* to program personnel.

Caution:

WATCH OUT FOR OVERLY AGGREGATED DATA!

Breakouts permit comparisons among groups (as discussed in Step 6). Breakouts also should be used to distinguish important groupings that have quite different outcomes from other groups. Subsequently, outcome analysis should track the progress being made separately for each group—to provide more meaningful information on what is happening.

What types of breakouts are likely to be useful to your program? Consider each of the categories shown in Exhibit 23 and discussed below. Each program needs to examine types of breakouts such as these to determine what breakouts will be useful for its particular outcome indicators.

Categories of Breakouts for Outcome Data

Geographical Breakouts

Geographical breakouts might, for example, be by state, region, congressional district, and/or zip code. Knowing the outcome of services in each geographical area will provide information to users about where service area outcomes are going well and where they are not.

Exhibit 23
Categories of Breakouts for Outcome Data

- By geographical location,
- By organizational unit/project,
- By customer characteristics,
- By degree of difficulty,
- By type of process or procedure used to deliver the service.

Organizational Unit/Project Breakouts

Those programs that support individual projects are likely to find separate outcome information on each individual project to be quite useful. The manager of each project should have outcome information that pertains to that manager's own area of responsibility. Outcome information that lumps together outcomes from more than one project is *not* likely to be very useful to managers of individual projects.

Similarly for programs that assist individual SEAs, LEAs, and even individual schools (such as the Star Schools Program, which supports some projects that focus on individual states, school districts, and/or schools), outcome data should preferably be collected and grouped so that the individual states, school districts, and schools would receive feedback. Doing this, however, can potentially be very expensive.

In such cases, the program will need to be satisfied with less than complete coverage of all individual units or much smaller sample sizes (and, thus, less precision) for individual units. One option is to include in the sample each unit that has large numbers of customers (such as schools or students) and combine as one category other units that have small numbers of customers.

For many Department programs, breakouts by school district and school characteristics will be important. For example, size, location, and demographic characteristics will be important breakouts for some outcome indicators (such as schools that fall within various ranges of students eligible for subsidized school lunches). Whether or not the school district or school has certain programmatic or organizational characteristics may be important for some Department programs. For example, some Department programs might believe it is important to assess whether end outcomes are related to such characteristics, as whether school districts or individual schools have introduced school-based management, have introduced new NCTM mathematics standards, or the extent to which districts are using charter or magnet schools.

Customer Characteristics Breakouts

Breakouts by categories of customers (or other forms of program workload) can be very useful in providing information to program personnel about the extent to which particular categories of customer services are achieving the desired outcomes and for which categories desired outcomes are not being achieved.

Some customer characteristics that may be relevant to your program include those listed shown in Exhibit 24, below. These apply particularly to outcome indicators about students and/or their families.

Exhibit 24
Examples of Customer Characteristics By Which Outcome Indicators Should be Broken Out (Disaggregated)

- Grade level,
- Age group,
- Gender,
- Race/ethnicity,
- School meal assistance eligibility / household income group,
- Household composition (such as size and number of children),
- Disability status,
- English-speaking capability,
- Other special status (such as migrant worker family).

For programs for which *organizations* are a category of customer such as SEAs and LEAs, breakouts might be by characteristics such as:

- Size (e.g., enrollment or number of teachers),
- Whether urban, rural, or suburban,
- Status of educational reform,
- Some other indicator of need relevant to the program's mission.

Breakout characteristics need to be tailored for each particular program.

Degree-of-Difficulty Breakouts

All programs are tasked with responsibilities that vary considerably in difficulty. More difficult workload means that the program can be expected to have a harder time achieving desired outcomes. This applies whether the program is assisting students, parents, teachers, states, or educational loan applications.

The degree of difficulty for various program components may vary and may differ from one reporting period to another. This can have significant effects on the outcomes.

What Difference Do Workload-Difficulty Breakouts Make?

Exhibit 25 indicates the importance of considering the difficulty factor. The exhibit illustrates that the outcome picture can change drastically if difficulty is considered. Unit #1 appears to have achieved better outcomes than Unit #2 when the data are examined in aggregate. However, when the incoming cases are broken out by level of difficulty, Unit #2 is found to have had higher success rates on *both* difficult and non-difficult cases.

This shows how easy it can be to jump to conclusions *if only aggregate data are reported. This is a major problem with overly aggregated data.* What happened? Unit #2 had a much higher proportion of difficult cases than Unit #1 (60% versus 20%). Because it is harder to help difficult cases than non-

difficult cases, Unit #2 looks worse if difficult and non-difficult cases are lumped together. When each difficulty group is considered by itself, however, Unit #2 is shown to have a higher percentage of successes for both types of cases.

Exhibit 25.
Workload (Client) Difficulty Breakout

	Unit #1	*Unit #2*
Total Clients	500	500
Number Helped	300	235
Percent Helped	60%	47%
Difficult Cases	100	300
Number Helped	0	75
Percent Helped	0%	25%
Non-Difficult	400	200
Number Helped	300	160
Percent Helped	75%	80%

Another Good Reason for Breaking Out Outcomes By "Difficulty"

Reporting breakouts by difficulty will eliminate the temptation for service delivery personnel (public personnel or private contractors) to *"cream" or "skim"*—that is, focus on easier-to-help customers.

To make its performance look good, an organization may be tempted to attract easier-to-help customers, while discouraging service to more difficult (and more expensive) customers.

How To Develop "Difficulty" Categories

The easiest and most practical way to develop categories of difficulty is to ask key persons who are fully familiar with your program to work as a committee to develop the difficulty categories. The committee should include persons who are knowledgeable about the details of program operation.

Committee members should be asked to establish the number of difficulty categories they believe are appropriate. This might be as few as two, but probably should be no more than five categories. The group's key task is to define each category in very specific terms based on customer characteristics for which information can be expected to be available to program personnel. The categories with their definitions should then be pilot-tested on a sample of cases (customers or other workload units such as educational loan applications). Persons who are expected to assign categories in the future to each workload unit should be asked to select the difficulty category for each case of the sample cases, based on the characteristics of each customer and by using the category definitions. The ratings of these individuals should then be compared to determine whether they are sufficiently comparable to allow for reliable ratings. If their ratings are not close enough to one another, the category definitions will need to be reworked, or the raters given more training.

A more technically sophisticated procedure is to use statistical analysis to relate available data on outcomes to the key characteristics. This procedure is considerably more complex and difficult—and probably not practical for small education programs.

Service Delivery Process or Procedure Breakouts

Relating outcomes to the type and magnitude of activities being supported by the program is likely to be of major interest to program managers. Most Department programs provide assistance to projects, each of which may use different types of activities and in various magnitudes. Thus, a program should consider breaking out outcome data by key characteristics of the projects being supported by the program.

For example, some parental assistance projects might focus on parent-school cooperation; others might focus on helping parents to be better able to help encourage their children in learning activities; others might focus on parent-teacher relationships; etc. The program could seek data for each project that it supports on: (a) the type and amount of each activity; and (b) the outcomes resulting from each project's efforts. From this information the program can produce combined outcome information for the projects grouped by type, and amount, of activity provided.

Such information can be very useful in distinguishing the more successful from less successful approaches.

To do this, however, the program will need to classify carefully the types/characteristics of its projects so that these data can be reliably collected on each project. The projects themselves, preferably, should be involved in the selection and definitions of these types/characteristics.

A Special Application, One Encouraging Innovation

The following variation of this type of breakout has seldom been used but can be quite useful to innovative program personnel.

The program can try out *new* procedures on some of its incoming workload, while continuing to use existing practices on the remainder of the incoming workload. The outcomes can then be compared between the existing and new practices—to see whether the new practice seems to be superior and should replace the existing practice.

Unlike the previous types of breakouts, the breakouts here are likely to apply only during the period of the "experimentation."

Exhibit 26 is a hypothetical example of a breakout of two procedures used for processing educational loan applications. Two outcome indicators are used to compare the procedures. (The data in the exhibit and the conclusions they might lead to are discussed in the next step on "comparisons.")

Exhibit 26.
Comparing Program Procedures for Processing Loan Applications: Computer vs. Manual Procedures

Procedure Used	Outcome Indicators	
	Error Rate	Time to Process the Applications: Percent Exceeding One Day
Computer Processing	9%	18%
Manual Processing	8%	35%

(Data are for three months of applications—about 250 applications processed by each procedure)

Steps to Use for Breakout Experiments

Here are steps you can use to undertake such experiments and provide outcome breakouts:

1. Identify the new practice and how it differs from the existing one.

2. Choose a procedure for selecting which incoming workload will be served using the new procedures and which will be served using the existing procedures.

   Preferably use some form of "randomization," even if only by flipping a coin. The purpose here is to select a *representative* sample of the workload for each procedure. You should seek to have approximately the same proportion of difficult workload in each of the comparison groups.

   Another approach is to assign every unit of incoming workload alternatively to each of the procedures. If the arrival of workload is essentially random, this would serve the same purpose as flipping coins or using random number tables.

3. As each unit of incoming workload is received, program personnel should assign it to one of the two groups by one of the above approaches.

4. Record which procedure was used for which item of workload.

5. Track the outcomes for the workload for each procedure over a period of time (a length of time the program believes is necessary to indicate fairly the outcomes of these procedures).

6. Tabulate the values on each outcome indicator for each of the two procedures.

7. Compare the findings and make future adjustments to program practices as appropriate. (You may want to drop the new procedure. Alternatively, you might find that you have not yet received a

clear enough picture from the outcome data, in which case you may want to continue the "experiment" longer.)

What Are Your Relevant Breakouts?

Many programs may have special characteristics of relevance to them that do not fall into one of the previous categories. The program's outcome measurement working group should examine each of its outcome indicators to determine what breakouts would likely help them identify where the desired outcomes are successful and where they are not.

Procedures for Selecting, and Then Collecting Data on, Breakouts

Selecting the Breakouts Wanted

The program should review the five breakout categories identified earlier in this step to determine which categories apply to the program. Preferably, the breakout categories would be selected by the program's performance measurement working group after obtaining input from program staff, other components of the Department, and customers. The various sources used to identify outcomes discussed under Step 2 should also be sources as to needed breakouts.

The program then needs to select the specific sub-categories for which outcome data breakouts are wanted. For example, by which specific grade level groupings should outcomes be sought? Grades K-6, 7-9, and 10-12? Or should data only be sought for specific individual grades, perhaps only those for which academic testing is commonly done? Or should the outcome data instead be sought by age groups, and, if so, which age groupings?

Another example: If intermediate outcome data on state progress in systemic reform area are being sought, are data needed on each state, or is it sufficient to report by region? Should states be grouped by size categories? If so, what should be the size ranges for each category?

Which sub-groups, and how many, that should be sought will be determined in part by: (a) the data collection procedures used; (b) the resources available to collect the data; and (c) the accuracy needed.

For example, if customer survey procedures are used to obtain data (such as from SEAs, LEAs, teachers, parents, or students), cost constraints may not permit large enough samples to provide sufficiently accurate data on more than a few sub-groups. For the level of accuracy (precision) the program believes it needs for each sub-group, the program will need to assure that enough respondents of each sub-group are included in the sample of those surveyed to provide that level of accuracy. Programs will likely need some help from statisticians to help them understand the tradeoffs among survey costs and data accuracy.

Collecting the Breakout Data

The program should decide how the data for the desired breakout information will be collected before data collection for the outcome indicators occurs. After data collection the breakout data may not be available or may require the program to go back and reconstruct the data, usually a very inefficient process. If demographic breakouts are to be sought and customer surveys are to be used, the survey questionnaire will likely need to include questions that provide at least some of the demographic information. (An exception may occur for surveys of states and school districts. The program may have sufficient information in its records to provide data on the characteristics of those organizations.) If the breakout data are expected to be obtained from program record information, the information collection process may need to be modified to capture the desired breakout data.

Step 6. Compare Findings to Benchmarks

When you have outcome data for a particular time period, how will you know whether that level of performance is good or bad? You need comparisons, that is, benchmarks, against which to compare current data. This step identifies the types of comparisons (benchmarks) useful for your program. This step is a major one in a program's analysis of the findings from its outcome measurement data collection. (This and other basic analysis steps are discussed further in Step 8.)

Some major types of comparisons (benchmarks) against which to compare actual performance for each reporting period are as follows. These benchmarks are discussed in detail below.

- Previous performance (improvement over time),
- Performance of similar units (e.g., benchmarking against the best),
- Outcomes for different client groups (e.g., benchmarking against the best),
- Pre-selected targets (a pre-selected standard),
- Different service delivery practices (use of comparison groups).

Categories of Benchmarks

Comparisons of Current to Previous Performance

This comparison will probably always be relevant and important (assuming data on previous performance are available for the outcome indicators). Comparisons of current to previous performance are applicable to all programs and are the most common type of comparison.

Current performance should be compared to that of previous reporting periods, whether the reporting periods occur monthly, quarterly, or annually. How frequently should the data for each indicator be reported? Frequent feedback will be more useful to program managers and staff than infrequent data. For Department reporting, annual reports may be sufficient. However, for field projects the Department's partners may need more frequent feedback, such as semi-annual or quarterly reports.

> *For example, the Star Schools Program is planning to report outcome measurement data annually. However, its various projects and school systems probably could use some of the information from the outcome measurement process if collected after each semester (such as data on the adequacy of distance learning materials).*

More frequent and timely feedback to program (and project) personnel is an important consideration. The down side of more frequent reporting, however, is its added time and cost. The program can, of course, choose its own frequency, indicating to state and local partners that more frequent feedback is their choice and responsibility.

Comparisons of the Performance of Similar Units

Comparisons among program units that provide essentially the same service (to approximately the same type of customers) are likely to be particularly useful to program personnel. Reporting such comparisons can also have motivational value for program personnel in each unit. Units might, for example, be different states, school districts, like schools, different universities, different projects or support organizations (such as Regional Labs or Comprehensive Centers).

The key concerns for such comparisons are that (a) the units are managed by different program personnel, and (b) the missions and types of customers are reasonably similar, so that comparisons will be meaningful and fair. For example, as the outcome data become available to the Star Schools Program, comparisons of project outcomes can be made. Such comparisons, however, will need to consider the types of clients and purposes sought by each project to make sure the comparisons make sense.

Caution:

Because of cost considerations, a program may be inclined to reduce the frequency of data collection for important indicators.

Do not reduce the frequency until you have considered low-cost data collection options such as those discussed in Step 4.

Exhibit 25 in Step 5, which illustrated breakouts by customer level of difficulty, provides an example of what can be done here. Comprehensive and fair comparisons of the performance of the two units in this exhibit require comparing the outcomes *separately* for the more difficult cases, for the less difficult cases, as well as in the aggregate.

Comparisons of Outcomes for Different Customer Groups

In Step 5, breakouts by various customer demographic characteristics were discussed. Once you have such breakouts, you can make comparisons among the categories.

Comparisons should be made to indicate whether the program appears to be more or less successful with certain categories of customer/workload than with others such as males compared to females, different age/grade groups, different racial/ethnic groups, different handicapped groups, and so on.

Such comparisons can focus the program's attention on customer/workload groups for which outcomes have been significantly lower on one or more outcome indicators.

For any breakout characteristics that you identified in Step 5, comparisons can and probably should be made.

Comparisons to Pre-Selected Targets

A highly useful management tool is to ask program managers to establish targets for each outcome indicator for the coming performance period(s). This is required of Federal programs by the Government Performance and Results Act. Federal programs have to set targets at the beginning of each year for the whole year—and later report to the President and Congress on the actual values compared to the targets.

Programs are likely to find it useful to set targets for *each reporting period* during the year, whatever the timespan of the reporting period.

Programs may also want to establish out-year targets, perhaps for five years into the future. These targets should be the same as those included in the program's strategic plan, if one exists. This will encourage longer range thinking by program personnel.

A program might want to include a range rather than a single value for some targets. For example, if the indicator is expressed as a percentage, the target might be the range encompassing the most likely percentage, plus or minus 5 percentage points.

Preferably, different targets should be set for each outcome indicator for each breakout category (see Step 5), especially for each level-of-difficulty category that the program has identified. This will make the comparisons much more useful and meaningful, and will provide fairer comparisons. It will also reduce the temptation for program personnel to concentrate on easier cases or to control the incoming workload mix in order to show high performance.

After experience is gained on individual outcome indicators, annual target levels of performance should be set for each outcome indicator. If an indicator is new, and past data are not available, note this and establish a plan and schedule for collection.

How should such targets be selected?

Targets can be set in many different ways. Here are some suggestions:

- Consider previous performance, almost always a major factor in determining targets.

- If the program has more than one unit that provides the same service (for approximately the same types of customers), consider using the performance level achieved by the *most successful* managerial unit as the target for all units.

  For example, in Exhibit 25 the program might select Unit #2's outcome rate for both "difficult" and "non-difficult" cases (that is, 25% and 80%, respectively, as the next year's target) — at least for internal Department reporting. For reports going outside the Department, to avoid overwhelming others with mountains of data, the aggregate "success rate" might be provided as the target. Such aggregated targets would need to be based on the program's best estimate of the likely mix of customers expected in the next reporting period.

  A more conservative option is to use the *average performance level* of all units. (If the program wants to be even more conservative, it could use the worst value as the target, to emphasize the need to achieve at least that minimum level of success.) Programs should avoid the temptation to underestimate targets in order to look good each year; program reviewers will eventually catch on.

- Consider the outcome levels achieved in the past for different customer (workload) categories. For example, as the target for all categories, use the highest or average level achieved for any one demographic category. If a program indicated successful outcomes for, say 53% of males and 48% of females, consider setting a future overall target of 53%—for each sex and in the aggregate. This is consistent with the Department's efforts to encourage high standards for all students.

- The target chosen should be feasible given the program's budget and staffing plan for the year. Programs may be pressured to keep the same target levels despite reduced budgets. This probably can be achieved up to a point, but at some level of cutback the program will be unable to provide the same level of outcomes. The program should be able to reflect this in the form of reduced target levels.

- Identify any new developments likely to occur during the coming period that may affect the program's ability to achieve its outcomes. Consider both internal and external factors. For example, legislative changes, whether policy or budget changes, recently made or expected to occur during the next reporting period might make it more or less difficult to achieve desired outcomes.

Comparisons of Different Service Delivery Practices

Programs periodically consider new, alternative practices. As discussed in Step 5, the outcome measurement process can be used to help them assess the results and outcomes of the new practices. The following types of new practices might be introduced:

- Different operating procedures
- Different technologies
- Different staffing arrangements
- Different policies
- Different amounts/levels of service provided to individual customers
- Different providers (such as private contractors)

A program can use its outcome measurement process to help it compare alternative policies, processes, or procedures to those currently being used. Two principal approaches to assessing these alternatives are:

- Introducing new practices across the board, or
- Introducing new practices into only part of the program operation, thus, running both the old and new practices side-by-side for a period of time.

In the first approach, outcome data can be used to *track changes in outcomes from before the change to the outcomes after the introduction of new program practices.* Data for periods *after* the introduction of the new practices should be compared to data from time periods *before* the introduction.

For example, suppose new automated equipment to help reduce response time to state requests for waivers is introduced in the middle of the first quarter of 1995 and is now available through the end of the third quarter of 1996. Exhibit 27 illustrates what the data can show.

Exhibit 27
Example of Comparison of Outcomes Before Vs. After Procedure Changes

	1994 Quarters				New Equipment Introduced in Middle of 1st Quarter, 1995 1995 Quarters				1996 Quarters				Average For 5 "Before" Quarters	Average For 6 "After" Quarters
	1st	2nd	3rd	4th	1st	2nd	3rd	4th	1st	2nd	3rd	4th		
Average Response Time (Days)	53	51	56	49	53	53	49	47	43	45	44	n/a	52.4	46.8

The data indicate that the new equipment appears to have improved response times. The equipment should be continued and perhaps be added to other parts of the system.

The second approach is to introduce a new policy, procedure, or process in part of the program while leaving the old practice in other parts. The example of computer versus manual education loan processing (see Exhibit 26 in Step 5) illustrates the outcome data that can be obtained. (Step 5 also describes the procedures that might be used to undertake such an experiment.)

The numbers in Exhibit 26 indicate that computer processing is considerably faster. However, error rates are slightly higher. A problem may be present with the error rate for computer processing (especially if the aim of the new procedures was to lower error rates substantially). The agency, based on this outcome information, might choose to review the errors made and attempt to correct any problems found. If and when the problems are corrected, the program might then choose to move completely to computer processing.

This second approach can help program managers test alternative program procedures and innovative ideas—without having to make a full commitment to any particular one. The approach can provide strong evidence that a procedure is preferable or not preferable.

Step 7. Pilot Test the Procedures

The program should test the new or substantially modified data collection procedures before full implementation. This is needed to identify and eliminate bugs in the process. Usually, it will be practical to test the procedures on only some segments of the program. For example, the program might test:

- Some of the projects funded by the program,
- Some of the locations that the program serves,
- The above for only part of the year, and
- Only those outcome indicators that are new or require substantial modifications to existing data collection procedures.

Preferably, the working group would oversee this test. The working group would work with program and project personnel to identify and alleviate problems identified during the test period. The group should also keep track of the staff time and out-of-pocket costs required for the new procedures during the test. An option is to seek resources from within the Department and the component in which the program resides. The Department has many strong analytical resources (such as the statistical and survey research capability of NCES) and often will be able to support a limited term pilot project.

After the test, the working group should make modifications to the various procedures to alleviate problems with the sampling plan, with questionnaires and other data collection instruments, and with definitions used for the individual outcome indicators. The group should estimate the annual costs to the program for obtaining the data (by extrapolating from the cost information obtained on the pilot effort). It should make recommendations regarding possible elimination of outcome indicators for which reasonably accurate data appears infeasible or excessively costly to obtain. The group might also identify additional outcome indicators that should be tracked.

The pilot test itself should be undertaken to approximate the conditions that will exist after full implementation begins. The testing team should look for the following problems:

- Definitions of the desired data that are unclear to those trying to collect the information.

- Missing data. For example, the list of teachers or students who went through the sponsored program might be incomplete so the test is not able to survey or obtain project records on these persons.

- The need for confidentiality and requirements for parental permission to survey students. This adds considerably to the effort required to obtain the data and reduces survey response rates.

The testing team will need to resolve such problems. Resolutions may involve correcting the problem, deciding to delete an outcome indicator, or accepting less accuracy than hoped.

Step 8. Analyze and Report Outcome Information

After the program has begun obtaining data on its outcome indicators, it needs to examine the data to identify where the program appears to be doing well and where not so well in achieving its mission and objectives. The program should not report the outcome data without examining it for problems.

Steps 5 and 6 discussed, respectively, identifying breakouts that the program should consider for each of its outcome indicators, and identifying comparisons that the program should consider making when examining its latest actual outcome data. Together these choices of breakouts and comparisons will likely be the major focus of a program's analysis effort after it receives the latest outcome information.

Major Analysis Elements

Program staff should examine each of the comparisons the program identified for each outcome indicator. These include comparisons of:

- The latest to previous performance. (For those outcome indicators for which data have been obtained for the first time, this option will not be available.)
- The outcomes for various units/projects for which breakout data have been provided, to determine whether some units/projects are not yielding the outcome levels that others have been able to achieve.
- The outcomes for different client groups for which breakout data have been provided, to determine whether the outcome levels for some groups are different than expected.
- Actual outcomes to the targets set by the program for each outcome indicator.
- The outcomes under the new procedures to the outcomes under the old procedures, when the program has experimented with new service procedures.

In each case, program staff should identify outcomes that indicate that a) the program has done considerably better than anticipated; or b) the program has done worse than anticipated. In either case, staff should attempt to identify why this has occurred. The program's analysis work should usually focus on these circumstances, particularly outcomes that show very large discrepancies or that occur for outcome indicators that the program believes are particularly important. Programs will most often focus on situations where outcomes were worse than anticipated, since that is where they are likely to be pressured by persons outside the program to do some explaining.

The breakout information should enable staff to narrow the search for reasons for unanticipated outcomes to the categories of customers and projects where these unanticipated levels of outcomes have occurred.

In some instances, the reasons for outcome shortfalls (or unexpectedly high outcomes) will be fairly obvious. Usually, however, they will not. Program staff can:

- Discuss the findings with key personnel in the program, Department, and in the field (with state or local education personnel, local project personnel, and contractors);

- Suggest corrective steps, even though the staff are not clear about what the problems were. (After outcome data reports are available, staff can assess whether the actions taken appear to have resolved the problems);

- Suggest that some form of in-depth evaluation be undertaken to attempt to identify causes and what correction should be taken; or

- Do nothing and hope that the unsatisfactory outcomes are an aberration and will correct themselves in the future. (The program may want to wait before taking any action to determine whether the problem is temporary or represents a trend.)

Reporting

How you report can be as important as *what* you report. This section identifies a number of possible reporting formats for your written performance reports. (Oral reporting is not addressed here.)

Outcome reports probably will be used to summarize the comparisons selected for the program (see Step 6). Below, is a list and brief description of a number of reporting formats.

As discussed in Step 3 on selection of outcome indicators, a program is likely to find it useful to track a relatively large number of outcome indicators for *internal* program use, but select a shorter list for *external* reporting. The Department's highest level officials, its budget office, and OMB, are likely to want a relatively short list of indicators. (Exhibit 18 presented in Step 3 illustrates a short list for external reporting for the Star Schools Program, extracted from the full list contained in Appendix 3.) The report formats described below can be used for both the internal or external lists of outcome indicators. They are just a sample of the many formats that can be constructed based on the special needs of the program. (Note: The data in all these exhibit formats are hypothetical.)

- ***Format #1, Exhibit 28, compares actual outcomes to targets for both the last and current reporting periods.*** It does this for each outcome indicator. This format is likely to be an important one for programs.

- ***Format #2, Exhibit 29, compares for a number of indicators the latest outcomes for various geographical locations. This format is also useful for making comparisons across any breakout categories identified by the program.*** For example, a program may want to illustrate comparisons across managerial units, individual projects, schools, school districts, or particular student characteristics. To do this, the program would change the column labels in Exhibit 29.

- ***Format #3, Exhibit 30, is similar to Format #1 but shows cumulative values for a year rather than for previous reporting periods.*** This format will only be useful for outcome measurement systems that provide data more frequently than once per year.

- ***Format #4, Exhibit 31, displays, by project, outcome data for one indicator (improved achievement) broken out by beginning student achievement level (or customer "difficulty") characteristic.*** This format is likely to be quite useful for internal reports, to show how each

project has performed relative to others. Displaying this information by key characteristics such as difficulty of the incoming workload will make the comparisons more fair and informative.

- ***Format #5, Exhibit 32, presents the responses from a customer survey of parents on a single outcome indicator—with the responses broken out by respondent characteristics displayed on one page.*** This is an internal format that enables program staff to identify which respondent (customer) characteristics show unusually positive or negative results for one particular outcome indicator. The format can be used to illustrate any indicator for which data on multiple customer (or program) characteristics can be obtained.

  Displaying the findings by respondent characteristics on the same page considerably eases the task of reviewers in spotting characteristics that have unusual findings. For example, the data in Exhibit 32 indicate that considerably poorer results have occurred in the central and northeast regions, for families with incomes below $30,000, and for nonwhite parents. These findings suggest that the program should consider giving special attention to parents with these characteristics. Data obtained in future years will indicate whether any such actions achieved improved outcomes for those groups.

The previous formats use *tables* to present the data. Other types of presentations can be used to supplement tables. A picture *is* often worth 1,000 words!

Other options include:

- Graphs for individual outcome indicators, to show the values of the indicator plotted against time, perhaps by quarter or year. This presentation is good for showing time trends.

- Bar charts, a special type of graph. (See Exhibits 33 and 34 for examples).

- Maps, which are a dramatic way to present geographic data. The geographical areas could be portions of states, showing the percentage of students or parents affected by particular educational reform efforts.

A special consideration for Federal programs is how to display results for use by the general public. It is beyond the scope of this manual to cover such presentation procedures. (With advanced computer processing many marvelously attractive displays can be generated.) Exhibit 35, however, illustrates a useful external presentation that Department programs should consider to encourage SEAs and LEAs to prepare for both internal and public use. It compares outcomes (on math test scores) for a particular school to average scores at each geographical level (school, district, and national), and includes the school's rank on math test scores in its school district. It also provides these comparisons for several years in the past. Similar formats may be useful for some Department programs.

A final suggestion on report formats is to highlight (e.g., circle) the outcome values that warrant attention.

Provide Explanatory Information For Both Analysis and Reporting

As an integral part of a program's outcome measurement process, the program should provide explanatory information with each outcome report. Such information gives the program the opportunity to explain significant outcomes, such as indicator values that were worse than expected. Here are some detailed suggestions on explanatory information:

1. The information can be qualitative, quantitative, or a combination of the two. Quantitative evidence is most desirable, but program staff judgments also can be presented.

2. Provide explanatory information when any of the comparisons show unexpected differences in outcome values, for example when:

   - The actual value for an outcome indicator deviates substantially from the target value (whether the actual value was much better or much worse),
   - The outcome values show major differences among operating units, categories of customers, or other workload units—and the reasons are not likely to be obvious to readers.

3. Consider both internal (to the government) and external explanatory factors. For example:

   - Internal factors affecting outcomes might include significant unexpected loss of program personnel or other resources during the reporting period.
   - External factors affecting outcomes might include unexpected changes in state or federal legislation.

4. Summarize and highlight the outcome information that particularly warrants attention in order to help users of the outcome reports be able to focus quickly on the important findings. Highlighting might be done by providing a special summary section along with the detailed data information.

5. One source of explanatory information is responses from customer surveys to questions that ask respondents why they gave a poor rating to particular characteristics. These responses should be examined by program personnel to determine if significant patterns of concerns are present that should be followed up.

6. Another important source of explanatory information are findings from recent program evaluations or other studies that provide relevant information about the outcome measurement data being reported. These studies are likely to be highly relevant.

   The findings of such studies are likely to provide a more in-depth look at the program and its achievements — and, in some instances, may actually supersede the outcome data collected as part of the routine outcome measurement process. At the very least, the findings from recent program evaluations on the program should be given prominence in the program's presentation of its outcomes. Such studies are likely to provide considerably more information as to the impacts of the program than is possible with outcome data alone.

7. The explanatory information should be attached as part of the report and should be brief and to the point.

Include in your explanations and in the highlights section any actions the program has taken, or plans to undertake, to correct problems identified in the outcome report.

Disseminate the Outcome Reports

Disseminate to everyone on the staff each outcome report shortly after it is available. This will encourage all program members to feel they are part of a team whose purpose is to produce outcomes that are as high as possible. As will be discussed in Step 9, the program manager preferably would hold a staff meeting on the outcome data. The discussion of the outcome data should help the program develop the explanatory material that should accompany the data.

The outcome report—including data, explanatory information, and the highlights/summary— should then be provided to offices outside the program. A question to consider is how much detail should be disseminated outside the program.

Each report probably should also be provided to the program's units/projects outside the Department, including customers and private organizations participating in delivery of the service. Here again, the program will need to decide what detail will be of interest to these groups. Breakout data by unit/project is likely to be of major interest to these units. Providing such information is also likely to encourage them to focus on future outcome improvement.

Document the Outcome Measurement Process

The program may find it useful to prepare "Outcome Indicator Specification Sheets" for each outcome indicator so that program personnel and others have a description of each indicator. Exhibit 36 illustrates a format for such documentation.

Another useful summary is to list the outcome indicators *grouped by data source*. One list would identify all the indicators whose data would be collected by *agency records*. Other lists would identify each indicator to be collected by *surveys of customers*, with a separate list for each category of customer (such as students, parents, teachers, SEA or LEA administrators). Another list would identify each indicator for which data are to be collected by *trained observer procedures*.

An important use for these lists of indicators is to provide a basis for developing (or modifying existing) data collection instruments, whether they be forms, survey questionnaires, or rating guides.

After completing its modifications based on the pilot test (discussed in Step 7), the program should document the key procedures to be used for future outcome data collection. This will help institutionalize the outcome measurement process. It also will help new program personnel become familiar with the procedures. The availability of documentation is also likely to raise the credibility of the data with persons outside the program—individuals who want assurance that stable, systematic procedures are being used.

Exhibit 28.
Reporting Format #1: Actual Outcomes vs. Targets
Using Indicators Developed by the Star Schools Working Group (Illustration Only)

Outcome Indicator	Last Period			Current Period		
	Target	Actual	Difference	Target	Actual	Difference
1. Percentage of districts classified as Title I that are participating in the program.	35%	25%	-10%	35%	30%	-5%
2. Percentage of teachers reporting that they use distance learning technology regularly in the classrooms.	20%	20%	0%	25%	30%	+5%
3. Percentage of teachers reporting increased student interest in subject area, attributable at least in part to the use of distance learning activities in their classes.	50%	60%	+10%	60%	55%	-5%

This format compares actual outcomes to targets for both the last and current reporting periods on each outcome indicator.

Exhibit 29
Reporting Format #2: Comparisons Across Geographical Locations (Illustration Only)

Outcome Indicator	Geographical Location			
	Eastern	Central	Mountain	Pacific
1. Percentage of K-12 schools participating in the Star Schools Program.	30%	15%	20%	35%
2. Number of students enrolled in Star Schools high school credit, college preparatory, or AP courses that had not been available previously.	1,500,000	600,000	850,000	1,950,000
3. Percentage of students reporting increased interest in school because of distance learning activities in their classes.	65%	90%	85%	75%

This format is useful for making comparisons across any breakout categories identified by the program. For example, a program may want to illustrate comparisons across managerial units, individual projects, schools, school districts or particular student characteristics.

Exhibit 30
Reporting Format #3: Comparisons of Actual Values vs. Target —Current and Cumulative: Parental Resource Centers (Illustration Only)

Outcome Indicator	Current Period (e.g. Quarter)		Cumulative for Year		Year's Target
	Target	Actual	Target	Actual	
1. Percentage of parents reporting knowledge or awareness of local parental resource center activities.	75%	70%	70%	65%	70%
2. Percentage of parents reporting that parental resource centers led to their taking a more active role in their child's development or education.	50%	65%	50%	60%	50%

This format is similar to Format #2 but shows cumulative values for a year rather than for previous reporting periods. This format will only be useful for outcome measure systems that provide data more frequently than once a year.

Exhibit 31
Reporting Format #4: Outcomes by Individual Project, by Achievement Level (Illustration Only)

Beginning Achievement Level	Percentage of Students Whose Achievement Had Improved from the Start of the School Year			
	Project 1	Project 2	Project 3	Total, All Projects
Low Achievement	58%	69%	61%	63%
Medium Achievement	35%	30%	54%	39%
High Achievement	52%	35%	56%	47%
Total	48%	44%	57%	50%

This format is useful for displaying outcome data for one indicator broken out by one demographic or customer "difficulty") characteristic—displayed for each project (or quarter). This format is likely to be useful for internal reports—to show how each project has performed relative to others. Displaying this information by key characteristics, such as difficulty of incoming workload, will make the comparisons more fair and informative.

Exhibit 32
Reporting Format #5: Breakouts of Responses to a Customer Survey By Demographic or Program Characteristics

Respondent Characteristics		Percent of Parents Reporting That Parent Resource Center Activities Led Them to Take a More Active Role in Their Children's Education				
		Not at all	A Little	Somewhat	Considerably	Total Responding
Sex and Race	White male	7	11	71	10	265
	White female	8	13	68	11	284
	Nonwhite male	11	30	53	6	36
	Nonwhite female	10	17	65	8	40
Age	18-34	13	16	58	13	272
	35-49	6	11	75	8	125
	50-64	3	11	80	5	105
	65 and over	6	11	74	9	123
Family Income:	less than $20,000	12	22	55	11	150
	$20,000-29,999	15	19	55	11	117
	$30,000-39,999	4	16	70	10	100
	$40,000-49,999	7	5	78	10	69
	$50,000-74,999	2	9	78	11	104
	$75,000 and over	3	2	88	7	85
Region:	1. Central	12	17	60	11	150
	2. Northeast	13	16	59	11	174
	3. Northwest	6	12	74	8	76
	4. Southeast	3	9	77	11	113
	5. Southwest	5	9	71	15	112
Total Characteristics		50	83	429	63	625

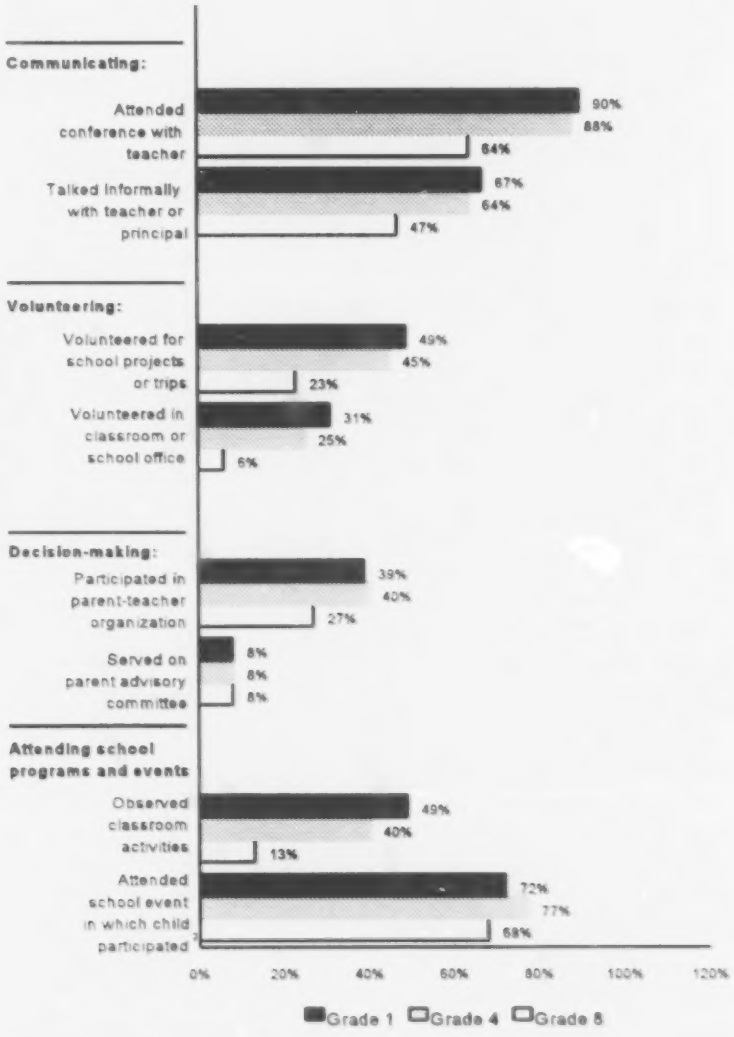
This is an internal format that enables program staff to identify which respondent (customer) characteristics show unusually positive or negative results on the particular outcome indicator.

Exhibit 33
Parent Participation in Specific School Activities

Direct Measure of the Goal:
School Leadership to Develop and Maintain Partnerships

Percentage of public school students whose parents reported that they[1] participated in the following activities at their child's school at least once during the current school year, 1992.

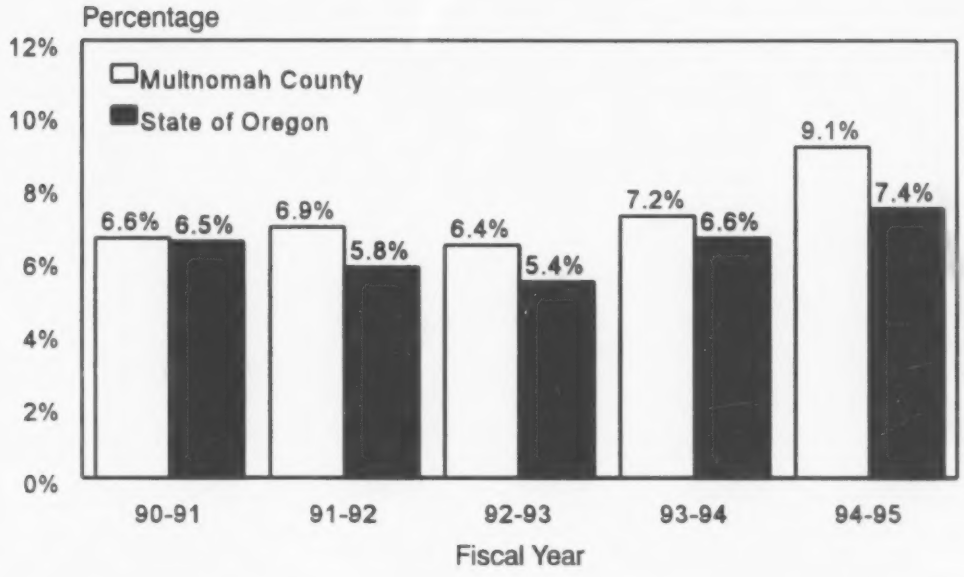
In 1992, parents of 8th grade students reported that they participated less frequently than the parents of 1st or 4th grade students in communicating, volunteering, decisionmaking, and school program activities



[1] Parent or another adult in household
[2] Such as a play, sporting event, or concert

Source: U.S. Department of Education, Planning and Evaluation Service, and Abt Associates, Inc. 1995.

Exhibit 34
High School Dropouts
Multnomah County, 1990-1995



Source: Oregon Department of Education, *Dropout Rates in Oregon High Schools, Salem, OR.*

Exhibit 35
Sample School Report Card for Parents, Taxpayers and the Business Community

School report card for 3rd graders	Source	1994	1995	1996	1997
Math scores for school	Test score	64	72	55	44
School district average	Test score	65	64	61	56
National Average	Test score	50	50	50	50
Schools rank on math scores (compared to others in school district)	Rank	57 out of 112 schools	38 out of 114 schools	73 out of 114 schools	NA

Number of students tested in 3rd grade: 58

Exhibit 36
Outcome Indicator Specification Sheet

Program: _____ Date: _____

1. Outcome Indicator

2. Category of Indicator (e.g., Intermediate or End Outcome)

3. Data Source

4. Breakouts of the Outcome Indicator That Are Needed

5. Frequency of Collection and Reporting

6. Who Is Responsible for Data Collection and Data Quality

Step 9. Use Outcome Information

This step suggests some ways to add to the usefulness of outcome information, and discusses a number of uses for performance information.

What Outcome Data Are Not Likely To Tell You

It is important to notify and remind your outcome report users that *outcome data do not tell how much of the observed outcomes was caused by the program and how much by other factors.* You might insert a note to this effect in every outcome report. Users of outcome information should recognize that outcome data tell you the "score", but not why the score is the way it is. They tell only whether you are "winning" or "losing."

Suggestions For Making Outcome Information More Useful

Earlier in this report we discussed ways to transform outcome data into useful information for program managers and other interested users. Four key elements were discussed in previous chapters: using breakouts (Step 5), comparing findings to benchmarks (Step 6), providing explanatory information (Step 8), and presenting the information in clear, understandable formats (Step 8). Following are three additional suggestions for making outcome information even more useful to your program.

Suggestion #1

Use the outcome information to trigger in-depth examinations of : a) why outcome problems exist and b) why the program appears successful in some situations and not in others.

To obtain information on WHY the "score" in a sporting event is the way it is, coaches need to examine the whys in more depth. Similarly, public program managers need to explore why the outcomes are the way they are.

A program might further examine the "whys" by undertaking or sponsoring in-depth studies such as systematic program evaluations. These can be costly in both time and effort. Most small programs are not likely to have the resources to do much of this. However, a program can also undertake more qualitative and less rigorous examinations.

Suggestion #2

Hold "How Are We Doing?" sessions after each outcome report.

A major use of the performance information is for the program manager to hold staff meetings shortly after getting the latest outcome data. At that session, the program manager and staff should address the following issues:

- Where have we done well? Why?
- Can we transfer the factors behind our success to other units of the program?
- Where are we not doing well? Why?
- What can we do to improve these results? The group should develop an action plan to implement these improvements.
- At later performance reviews, the program personnel should review the extent to which the desired outcomes have been achieved *after* the improvements were made.

Suggestion #3

Provide incentives for encouraging high program performance.

Ways to encourage high program performance include:

- Giving recognition awards for projects or units whose outcomes have reached high levels or whose outcomes have improved substantially.
- Having managers review and discuss outcomes in program meetings.
- Having the program manager and staff hold regular (e.g., quarterly or semi-annual) reviews after each outcome report (see above discussion of "How Are We Doing?" sessions).
- Using outcome information to develop and justify the program's budget proposals.
- Considering outcome information in performance appraisals, perhaps with a greater focus on appraisals of program staff as a group.
- Providing more flexibility to personnel when outcome levels meet the program's targets.

Uses for Program Outcome Information

The primary use of outcome information is to provide regular feedback to program personnel to encourage improvements that lead to improved outcomes.

Various specific uses of outcome information have been described throughout this report. The information can be used to help:

- Motivate employees to continually improve both services and the outcomes of these services.
- Track whether actions taken in the past have led to improved outcomes.
- Identify where problems exist and where action is needed to improve program outcomes.
- Communicate with and inform customers and the public at large about the extent to which education-related progress is being made.
- Support annual and long-range planning.
- Prepare budgets and justify resource allocation.
- Serve as a basis for setting annual targets.
- Identify needs for technical assistance and training for agency personnel.
- Determine the program's future evaluation agenda.

- Focus grantees and contractors on the mission of achieving results—for instance, by using "performance contracting" in which outcome targets are included in grants or contracts.

Special Note on Use of Outcome Information for Grants and Contracts

If the program provides grants or contracts for services to customers, the outcomes of those services should be tracked, especially if this work is expected to be done on an ongoing basis, rather than as a one-time effort.

The outcome indicators identified for the program as a whole should also be considered for use with grantees and contractors. *Outcome measurement preferably should be undertaken by each grantee or contractor,* both to help the national program prepare nationwide reports and to help improve the services for which each grantee or contractor is responsible.

Targets might be included in grant agreements and contracts on each outcome indicator that pertains to the relevant customers or workload. Such targets should be developed and agreed on jointly by the program and the grantee or contractor.

This process could eventually be extended to include incentives for successful outcomes in the grants or contracts. A combination of rewards and penalties could be included in contracts such as rewards for contractors who meet or exceed the target levels and penalties for those who do not.

Quality Control

Once the basic outcome measurement is in place, the program will need to assure itself and outsiders that the outcome data continue to be of good quality. It may be that quality control reviews are best left to a central-level office. The program, however, should review its data collection procedures on a regular basis to check that the intended procedures are being used and used properly. In the early years of outcome measurement, the program should probably review the procedures each year.

Some Special Problems

By now it should be apparent that the process of outcome measurement demands considerable effort. Program managers should expect to face some problems as they attempt to implement and sustain a significant outcome measurement process. A few of these problems are described here.

1. Managers and staff will likely not have had much if any exposure to or training in outcome measurement. Convenient training opportunities in outcome measurement are likely to be hard to obtain. Small programs in particular may have trouble finding the funds and/or staff for the training and technical assistance. Managers will probably need to obtain outside assistance from analytical offices within the Department or from an outside consultant. However, the program should avoid delegating full responsibility for developing the outcome measurement process to outsiders. One objective of such outside assistance should be to develop within program managers and staff a practical understanding of and commitment to outcome measurement.

2. Changes in legislative and Department priorities, along with periodic funding uncertainties, can easily discourage development of a process that might be significantly affected by these changes. These changes will, however, not generally substantially change the mission of the

program. A well-constructed outcome measurement process should contain most of the elements that are likely to be needed after such changes do occur. The most likely changes would be intermediate outcome indicators, since these are most affected by changes to the service delivery approach used.

3. Many if not most education programs sponsored by the Department are ultimately delivered by other levels of government or by the private sector. Much of the data for outcome measurement need to be provided by, or at least with the cooperation of, these other levels and sectors. State and local organizations usually need the same type of outcome information, though probably in more detail, such as on a school-by-school or classroom-by-classroom basis.

This means that many if not most programs should consider ways to develop data collection partnerships with these other organizations. Such partnerships and the resulting cooperation can substantially increase the logistics and time requirements for the outcome measurement development to assure that all interests are met. However, this joint effort should pay off in the long run by yielding a much better and smoother running outcome measurement process.

Summary of Key Issues

The primary new thrust of the GPRA legislation, of the Executive Order on Setting Customer Service Standards (E.O. 12862), and of the Department's own emerging performance measurement process is on indicators of outcomes. This guide provides recommendations for development of an outcome measurement process for individual educational programs. The relationship to these other types of performance indicators has been discussed, but this guide focuses on outcome indicators.

Outcome measurement should start with a program identifying its mission and objectives—but be cast in terms of the results that the program ultimately seeks to achieve. A sound comprehensive statement of mission and objectives is the starting point from which specific outcome indicators should emerge. Here are our summary suggestions:

How Should the Outcome Indicators Be Chosen and By Whom?

The outcome indicators for specific programs should, in general, be selected by program managers and their staff. However, the indicators should:

- Be compatible with, and support, the Department's and POC's strategic plan and the Secretary's performance agreement with the President.
- Be selected by program personnel with the assistance of other Department representatives and developed from a variety of sources.
- Encompass the concerns of the program's customers. Customers include the Department's partners (such as state education agencies), and ultimately students, their families, businesses (because of their need for an educated workforce), and taxpayers. Programs should, wherever feasible, seek direct input from their customers as part of their selection of outcome indicators. (This might be done, for example, by holding focus groups with each category of customer.)

How Frequently Should Outcome Information Be Reported?

To help program managers better manage their programs, outcome measurements probably are needed at least on a semi-annual, if not quarterly, basis. However, data for some indicators that are not expected to change frequently or are too expensive to collect frequently, might be collected less often—say annually or every other year. For budgeting purposes and for reporting to Congress, the President, and the public, the Department usually will need only annual reports.

How Will Data Be Collected? How Much Additional Work Will Outcome Measurement Mean for the Program?

This will depend on the specific outcome indicators and data collection procedures selected by the program office.

Establishment of the outcome measurement process usually requires significant effort by program personnel. Once the data collection and reporting procedures have been identified and piloted for feasibility, *on going implementation* should be planned so as not to be overly burdensome on program staff or on grantees.

For some outcome indicators, existing data sources will be sufficient. Other indicators, however, inevitably will require new, or at least major modifications, to existing data collection procedures. For example, regularly repeated surveys of customers will require new procedures for most Department programs.

The program may need Department assistance on customer surveys to alleviate the burden of surveys on individual programs, especially the routine, administrative aspects of surveys. To save cost, perhaps a small number of questions could be added to one of the major surveys already administered by NCES. Or, perhaps, a small number of small programs that require information from the same target population might collaborate to administer one survey instrument to serve the respective needs of each program. The program, however, may have funds to use outside contractors to administer regular surveys. While potentially costly, this will ease the program staff's burden of administering the surveys.

A Major Component of an Outcome Measurement Process: The Need to Disaggregate Outcome Data

To make the outcome measurement most useful to program personnel, the data for many, if not most, performance indicators should be broken out (disaggregated) by key characteristics likely to affect service quality and outcomes. Such breakouts as the following are likely to be important for the program to obtain a much better perspective on where a program is leading to successful outcomes and where it is not:

- By "project" (where the program sponsors a number of similar efforts) or category of project (such as training, dissemination, or technical assistance, projects);
- By geographical location, such as particular states;
- By demographic characteristics of LEA and student populations (such as age/grade level), income group (e.g., eligibility for subsidized school lunches), race/ethnicity, presence of learning disability, and whether the schools are urban, suburban or rural;
- By any special characteristics that are expected to affect substantially the difficulty of the program in obtaining desired outcomes. For example, a parental assistance program might want to tabulate outcomes separately for communities with various proportions of parents who have literacy problems, and are single parents; and
- By type of procedure or practices used by the program, by states, and/or by school districts. This will be particularly useful if the program experiments with different approaches — to help the program assess the outcomes of alternative practices, policies, or procedures.

How Should Targets Be Set Each Year for the Outcome Indicators?

The Government Performance and Results Act *requires* agencies to set *annual* targets for each outcome indicator and to report on the actual values at the end of each year (beginning FY 1999 with first target submissions due September 1997). However, the program can take advantage of this requirement and use the target-setting component to help manage the program. Program managers should consider establishing quarterly or semi-annual targets (or whatever internal interval is chosen for its performance reports—considering the needs of individual projects). Targets should be based, in part, on the expected timing of accomplishments and seasonal factors. Past performance, once such

data are available, will be a major factor in setting targets. The program's budget, which provides the resource constraints on the program, will be another major determinant of expected results. Funding uncertainties will inevitably cause uncertainty about targets. Assumptions as to funding (and possible other legislative changes) should be identified when submitting the targets.

What Analyses and Report Formats Should Be Used?

Programs offices will likely want to develop their own analysis plan. However, some common features are likely. These include such elements as the following:

- Examination of the breakout information to identify conditions under which desired outcomes are and are not occurring.

- Comparisons of the latest data: a) to previous reporting periods; b) among various customer demographic groups; c) among different procedures used by the program; and d) to targets set at the beginning of the year. Such comparisons should be presented in the program's outcome reports. This will greatly increase the information content and its usefulness for both program staff and readers of the reports at all levels in the Department.

- The program staff should examine in greater depth *why* key findings in the particular report have occurred (such as significant shortfalls or significant successes indicated by the outcome data). Explanations should be included with outcome reports.

- Outcome reports should be user-friendly! Data should be fully and accurately labeled. The report should avoid being overcrowded, with too much information jammed in. Report highlights should be provided along with the more detailed data.

Another Major Component of the Outcome Measurement Process: Inclusion of Explanatory Information

Programs should provide relevant *explanatory information* along with the outcome data. (Remember that outcome data by themselves do not tell *why* the measured outcome values are the way they are.) Such information is likely to be especially important when the data show unexpectedly poor, or good, results. This explanatory information might be in the form of other data or be qualitative judgments as to why the particular results occurred. Such information should enable users of the information to obtain a much better perspective on performance. This feature also gives program personnel the opportunity to give their views as to what happened and why. For example, problems during the reporting period may have occurred because of uncontrollable outside factors, such as legislative changes or problems in state or local implementation of needed actions, or because of internal factors such as the loss of key program staff.

What Happens If the Performance News Is Bad?

Almost certainly, each outcome report will identify some outcome indicators with shortfalls relative to what had been expected (such as compared to the targets for the reporting period). Any outcome measurement system is, in fact, intended to surface problems of below-par outcomes so that those who can do something about them are alerted and, after corrective actions are taken, are able to assess whether the actions have produced desired results.

Program managers and their staff should attempt to identify the reasons for the bad news and take steps, to the extent possible, to correct problems. The Department policy is not automatically to blame program managers and their staff for shortfalls. Many factors can contribute to poorer-than-expected results for the reporting period. As noted earlier, program managers are advised to include with their outcome reports explanations as to why any significantly lower outcomes occurred as well as the program's plan for actions to correct the problem.

How Does Outcome Measurement Relate to Formal Program Evaluations?

Many Department programs sponsor occasional program evaluations. These can provide substantial information on program performance. They have the advantage that they seek direct information on program *impacts*, that is, the extent to which the programs evaluated actually contributed to the measured outcomes.

Formal program evaluations, however, can only be done very infrequently on most programs, and, thus, provide neither the regular feedback on program progress needed to help managers manage, nor the annual feedback required by the President and Congress.

The manager of the program sponsoring an in-depth program evaluation should determine what information from the evaluation should be included in the program's current outcome report. The major findings from evaluations completed during a reporting period should be considered for inclusion in the outcome report. Such information can provide a highly useful complement to regularly collected outcome information from the program's ongoing outcome measurement process.

Furthermore, the availability of outcome data can help program evaluations by providing data on outcomes for current and past years. The regularly collected outcome data should also help the program determine its *future* program evaluation needs by identifying problem areas that the outcome measurement data indicate need attention.

What Are the Specific Uses of the Outcome Information?

As noted in the introductory chapter, outcome measurement has four basic uses:

- First, and foremost, outcome information should help program managers and their staff track how their programs are doing and with that information help guide improvement efforts. The information should for example, indicate where, when, and under what conditions, outcomes appear to be satisfactory and not satisfactory.

- Second, outcome information can be useful for developing and justifying budgets and for formulating recommendations as to needed legislation and policy.

- Third, outcome information is used by the President, Congress, and Department officials in helping to achieve accountability of programs for program quality and outcomes.

- Fourth, outcome information can be used by the program to help communicate with and inform customers and the public at large as to the extent to which education-related progress is being made.

Some more specific uses for the information provided by outcome measurement that program managers should consider include:

- Use the reports as a basis for obtaining staff input into programmatic and policy decisions for improving outcomes. Shortly after each outcome report is available, hold a "How Are We Doing Session?" with program staff. Ask the group to identify the successes and how these might be maintained or transferred elsewhere. Discuss the shortfalls and what might be done to overcome these in the future. Ask the staff to develop an action plan based on the actions suggested at the meeting. In later "How Are We Doing Sessions?" assess whether the actions taken have helped and whether they should be continued or modified.

- Use the findings to help develop annual budgets and, subsequently, justifications for the budget proposals.

- At legislation development or reauthorization time, use the findings to help with these activities, including responding to legislative (and media) queries about the program.

- Use the findings to help track progress in achieving the objectives in your strategic plan.

- Include the outcome indicators in agreements with grantees and contractors, where appropriate. Use the findings to encourage grantees and contractors to focus on service quality and results. Report the findings on all projects to each project to encourage improved performance in those grantee/contractor projects whose performance has fallen behind.

- Compare the outcome data to data on costs to provide indicators of efficiency.

- Use the findings to help the program develop its annual program plan. Problem areas detected by the outcome reports, such as particular projects or demographic groups for whom the outcomes were not as good as expected might, for example, become the focus for an action agenda.

- Use the findings to help determine the program's future evaluation agenda. For example, programs or projects whose performance was not as satisfactory as expected, or was particularly unclear, might be good candidates for future in-depth evaluations.

If the information from the emerging outcome measurement process is not useful to the program managers and their personnel to help them improve the program, the process will have been a failure. A useful outcome measurement process requires not only technical soundness but also the efforts of program personnel to make it useful for improving program outcomes.

What Are the Limitations of Outcome Measurement Information?

Major limitations are the following:

- Outcome data obtained will not usually tell the *impact* of the program on the measured outcomes. Data on outcomes do not tell *why* the results are as they are. Usually other factors outside the control of the program (and probably the Department as a whole) contribute to the results. This particularly applies to indicators that attempt to measure desired end outcomes. This means that worse-than-expected, or better-than-expected, outcomes should not be the occasion for automatic blame-setting, or automatic praising, of the program. Additional examination is needed to assess causes for shortfalls, or for better than expected levels, of desired performance.

- The state-of-the-art of outcome measurement is limited. Perfect measurement and complete coverage of all relevant program outcomes should not be expected by anyone. The objective of outcome measurement is to provide *better* information on program quality and outcomes, not *perfect* information.

How Does the Outcome Measurement Process Relate to Strategic Planning?

The Department and POC strategic plans should provide starting points for identifying outcome indicators that the program should track. Preferably, each program would have its own strategic plan, one complementing both the Department and POC plans. Those strategic plans should contain statements of mission and objectives. They should be the starting point for the program's development of its outcome measurement effort.

The outcome indicators developed for regular program outcome measurement should be compatible with, and support, the outcome indicators contained in the strategic plans of the Department, the Principal Operating Components, and the program.

The indicators included in strategic plans, however, are often considerably broader than those the program needs for its own managerial uses. The program will need to develop more detailed outcome indicators specific to its own needs. In effect, the regular outcome measurement process provides the operational, annual indicators of progress toward the objectives in the strategic plans. At least some of the indicators included in the outcome measurement process should provide periodic progress information to indicate how well the strategic plans' strategies are working.

A Final Word

Implementation of a sound, valid, and useful outcome measurement process requires hard work. Unless this effort produces information that is useful to the program, the effort will have been wasted. Regularly-obtained outcome information can potentially play an important if not vital role in program management and improvement. Whether or not the outcome measurement process will prove useful, to a considerable extent, will depend on the continuing support provided by the program manager and the Department.

References

Customer Surveys

a. *Your Guide to Measuring Client Satisfaction*. Office of the Comptroller General, Evaluation and Audit Branch, Communications and Coordination Directorate (Ottawa, Canada) 1992.

b. *Research Methods for Social Work*. (2nd ed.) Rubin, Allen, and E. Babbie. (Pacific Grove, Ca. Brooks/Cole), 1992. See Chapter 7, Constructing Measurement Instruments, and Chapter 8, The Logic of Sampling.

c. *Measuring Client Satisfaction: Developing and Implementing Good Client Satisfaction Measurement and Monitoring Practices*, Office of the Comptroller General, Evaluation and Audit Branch (Ottawa, Canada), August 1992.

d. Citizen Surveys: *A Special Report on Designing, Conducting, and Understanding Citizen Surveys*. T.I. Miller and M.A. Miller. International City Management Association (Washington, D.C.) 1991.

e. *Telephone Survey Methods, Sampling Selection and Supervision*. (2nd ed.) P.J.Lavrakas, SAGE Publications (Beverly Hills, Ca.), 1987.

f. *How to Conduct a Citizen Survey*. Planning Advisory Service Report Number 40, T. van Houten and H. Hatry. The American Planning Association (Chicago, Il.), 1987.

g. *Survey Questions. Handcrafting the Standardized Questionnaire*. Jean M. Converse and Stanley Presser. SAGE Publications (Beverly Hills, Ca.), 1986.

h. *How to Conduct Surveys: A Step-by-Step Guide*. A. Fink and J. Kosecoff. SAGE Publications (Beverly Hills, Ca.) 1985.

I. *Asking Questions: A Practical Guide to Questionnaire Design*. S. Sudman and N.M. Bradburn. Jossey-Bass Publishers (San Francisco, Ca.), 1982 .

j. *How Effective Are Your Community Services? Procedures for Measuring Their Quality*. The Urban Institute and International City/County Management Association. (Washington, D.C.) Second Edition, 1992—specifically Chapters 13 and 14, which focus on survey procedures, and Appendix 11, which provides guidelines for contracting for surveys.

Trained Observer Procedures

a. "Use of Ratings by Trained Observers," John M. Greiner. Chapter 10 in *Handbook of Practical Program Evaluation*. Jossey-Bass Publishers (San Francisco, Ca.), 1994.

b. *How Effective Are Your Community Services? Procedures for Measuring Their Quality*. The Urban Institute and International City/County Management Association (Washington, D.C.) Second Edition, 1992—specifically Chapter 12 and Appendix 10.

Performance Measurement and Reporting

a. *Organizational Performance and Measurement in the Public Sector*, edited by Arie Halachmi and Geert Bouckaert. Quorum Books (Westport, CT),1996.

b. *Accountability for Performance: Measurement and Monitoring in Local Government*, edited by David N. Ammons. International City/County Management Association (Washington, D.C.), 1995.

c. "Eleven Ways to Make Performance Measurement More Useful to Public Managers." H. Hatry, C. Gerhart, and M. Marshall, *Public Management*, Sept. 1994.

d. "The Case for Performance Monitoring." Wholey and Hatry, in *Public Administration Review*, vol. 52, no. 6, American Society for Public Administration (Washington, D.C.), Nov./Dec. 1992.

e. *Say It With Charts: The Executive's Guide to Successful Presentations in the 1990's*. Second Edition, G. Zelazny. Business One (Irwin, Illinois), 1991.

f. *Service Efforts and Accomplishments Reporting: Its Time Has Come: Elementary and Secondary Education*, Governmental Accounting Standards Board (Norwalk, CT.) 1989.

g. *Misused Statistics: Straight Talk for Twisted Numbers*, A.J. Jaffe and N.F. Spirer, Marcel Dekker, Inc. (New York), 1987.

h. *Using Performance Measurement in Local Government*, Paul D. Epstein, Van Nostrand Reinhold Co. (New York, N.Y.) 1984.

I. *The Accountable Agency*, Reginald K. Carter, SAGE Publications (Beverly Hills, Ca.) 1983.

Appendixes

1. **Sample Star Schools Teacher Survey**
2. **Sample Star Schools Student Survey**
3. **Star Schools Program Outcome Indicators**
4. **Trained Observer Procedures**

Appendix 1: Sample Star Schools Program Teacher Survey[b]

1. How long have you been teaching? Please check one.

   ___ Less than 1 year
   ___ At least 1 year, but less than 2 years
   ___ At least 2 years, but less than 4 years
   ___ At least 4 years, but less than 6 years
   ___ More than 6 years

2. What subject areas did you teach this past school year in which you used or attempted to use distance learning? Please check all that apply.

   ___ English/language arts
   ___ Math
   ___ Science
   ___ Social Studies
   ___ History
   ___ Geography
   ___ Other (please specify) _____

3. What grade levels did you teach using distance learning this past school year? Please circle all that apply.

   Pre-K K 1 2 3 4 5 6 7 8 9 10 11 12

4. How long have you been a [insert appropriate local name for distance learning effort] teacher? Please check one.

   ___ Less than 1 year
   ___ At least 1 year, but less than 2 years
   ___ At least 2 years, but less than 4 years
   ___ At least 4 years, but less than 6 years
   ___ More than 6 years

5. Approximately how many hours did you use [insert appropriate local name for distance learning effort] distance learning materials over this past year in your courses? Please check one. *[for Indicator #18, and maybe #16 in Appendix 3]*

   ___ 1 to 5 hours
   ___ 6 to 10 hours
   ___ 11 to 15 hours
   ___ 16 to 20 hours
   ___ 21 or more hours

[b] Many items in this instrument are based on a survey of classroom coordinators designed by the Pacific Star Schools Partnership / ESD 101 - STEP Star Program.

6. How would you rate the effectiveness of the [insert appropriate local name for distance learning effort] learning materials (e.g., workbooks, hands-on kits, electronic field trips, video tapes, homework assignments) provided to students? Please circle a number on the following scale. *[for Indicator #15]*

Very effective 5 4 3 2 1 Not effective

7. For the past school year, how satisfied were you with the *instructional support materials* which [insert appropriate local name for distance learning effort] provided to you. Please circle the number on the five-point scale that best applies for each of the dimensions shown below. *[for Indicator #20]*

	Very Satisfied				Not Satisfied
a. Understandability	5	4	3	2	1
b. Helpfulness	5	4	3	2	1
c. Adequacy/Completeness	5	4	3	2	1

8. Was the distance learning *equipment* provided to your school (please check one for each item) *[for Indicator #19]*:

a. Made available in a timely manner? ___Yes ___ No ___ Not sure
b. Working properly? ___ Yes ___ No ___ Not sure
c. Adequate for its purposes? ___ Yes ___ No ___ Not sure

If you checked "No" to any of the above, please describe the problem.

9. During the past school year, have you made any significant improvements in your teaching practices because of the [insert appropriate local name for distance learning effort] program? *[for Indicator #14]*

____ Yes ____ No

If yes, please briefly describe the improvement practice(s) below.

10. Overall, how would you rate the helpfulness of [insert appropriate local name for distance learning effort] to you in your teaching? Please check one.

___ Not at all helpful
___ A little helpful
___ Somewhat helpful
___ Considerably helpful

11. To what extent do you believe your students were able to improve their learning because of the use of [insert appropriate local name for distance learning effort]? Please check one. *[for Indicator #32a]*

___ Not at all
___ A little
___ Somewhat
___ Considerably

12. To what extent do you believe that at least in part because of distance learning activities your students (Circle the appropriate number for each response) *[for Indicators #32b,c,d]*:

	Not at all	A little	Somewhat	Considerably
a. Had increased interest in the subject area	1	2	3	4
b. Increased their attendance	1	2	3	4
c. Improved critical thinking and problem solving	1	2	3	4

13 a. During the past school year, did you participate in any professional development activities?

___ Yes ___ No

b. If "yes," were any of the professional development activities conducted by distance delivery media (e.g., telecast, teleconference, the Internet)? *[for Indicator #21]*

___ Yes ___ No

c. If "yes," how would you rate the following characteristics of the distance learning media used in these professional development activities? Please circle a number on the following scale *[for Indicator #23]*.

	Very Good				Very Poor
a. Understandability	5	4	3	2	1
b. Relevance	5	4	3	2	1
c. Overall Effectiveness	5	4	3	2	1

If you circled "1" or "2" for any characteristic, please describe the problems you saw.

14. What suggestions do you have for improving the [insert appropriate name for local distance learning effort] in the future?

Appendix 2: Sample Star Schools Program Student Survey[c]

I. Background Information

1. What is your current grade?

2. In what subject areas during the past school year did you have the opportunity to use distance learning materials and equipment? (Please check all that apply.)

   ___ English/language arts
   ___ Math
   ___ Science
   ___ Social Studies
   ___ History
   ___ Geography
   ___ Other (please specify) _____

3. Was this your first experience with distance learning? Yes _____ No _____
   If no, how many previous classes have you taken in which distance learning was used?

4. Which of the following best describes the type of distance learning activity in which you participated during the past school year?

   A. Full course
   B. Curriculum module
   C. Field trip or teleconference (single event)
   D. Computer networking
   E. Did not participate in distance learning this year

II. Distance Learning Courses and Activities

5. Approximately how many hours per week did you spend on distance learning activities for each of the following items:

   a. Classroom activities? _____ hours per week.
   b. Homework assignments? _____ hours per week.

6. To what extent do you think that the Star Schools activities added significantly to the quality of the course? (*for Indicator #29*)

   ____ Not at all
   ____ A little

---

[c] In these questions, substitute for "distance learning" any terms that you believe to be clearer to students in the school.

____ Somewhat
____ Considerably

7. To what extent were you able to apply the information learned or acquired through distance learning to your other courses, activities, or interests? *(for Indicator # 29)*

____ Not at all
____ A little
____ Somewhat
____ Considerably

8. Do you think distance learning courses (or courses that use distance learning activities) are better than courses not using distance learning methods? *(for Indicator #31)*

____ They were considerably worse.
____ They were somewhat worse.
____ They were somewhat better.
____ They were considerably better.
____ Not sure.

Please tell us why.

9. Do you feel that your participation in distance learning courses or activities affected your interest in the subjects where the technology was used? *(for Indicator # 30)*

____ Increased my interest considerably
____ Increased my interest somewhat
____ No effect
____ Decreased my interest somewhat
____ Decreased my interest considerably

Please tell us why.

10. Would you take another distance learning class or participate in future distance learning activities?

____ Yes ____ No ____ Not sure

Why or why not?

III. Distance Learning Equipment and Media

11. To what extent was the distance learning equipment *available* in your school?

____ Always or almost always available
____ Usually available
____ Usually unavailable
____ Always or almost always unavailable

12. To what extent was the distance learning equipment *working properly* in your school?

____ Always or almost always working properly
____ Usually working properly
____ Usually not working properly
____ Always or almost always not working properly

13. Was assistance made available on how to use the equipment?

___ Yes ___ No

14. How would you rate the usefulness of the distance learning materials (e.g., workbooks, hands-on kits, electronic field trips, videotapes, homework assignments) used in your course or distance learning activity?

a. No materials used
b. Not at all useful
c. A little useful
d. Somewhat useful
e. Very useful

THANK YOU!

Appendix 3: Proposed Outcome Indicators for Star Schools Program Annual Performance Measurement[a]

Mission: To improve student learning and employability by providing access to and improving instruction in a wide range of subjects through the use of distance learning technologies.[b]

OBJECTIVE #1: ACCESS
Reach underserved learners of all ages throughout the U.S. and its affiliated territories, and expand instruction in core subject areas as well as literacy skills and vocational education.

OUTCOME	INDICATOR [c]	1996-97 TARGETS	DATA SOURCE
A. The program reaches different types of learning communities.	1. Number and percentage of K-12 schools (of all schools in the targeted areas) participating in the program,[d] by school characteristics such as grade levels, rural, urban, suburban, state. (I)[e]	At least 20 percent of all participating schools in the target areas for current projects are urban.[f]	School records
	2. Number of other settings, by characteristic, where learners participate in the program. (I)	Number of other settings increases by 25 percent from ____ (?)[f]	Project records
	3. Number and percentage of districts classified as Title I who are participating in the program. (I)	Percentage of Title I districts participating increases by 10 percent from ____ (?)[f]	Project records
	4. Number and percentage of (a) K-12 students (ethnicity, age/grade level, gender); (b) teachers (ethnicity, level of experience); and (c) others participating in the program. (I)		School records, district records
	5. Total learner program hours: (a) satellite, (b) modules, (c) other. (I)		School records, district records
B. Underserved learners of all needy categories participate in the program.	6. Number and percentage of K-12 students who are disadvantaged, LEP, or have disabilities, who are participating in the program. (I)[e]		School records, district records
	Number, by type, of program participants beyond K-12, such as those eligible for adult basic education. (I)		School records, district records
	8. Examples of program activities that reach beyond the intended audience or to other aspects of the school, district, or community. (This is a qualitative indicator.)		Project records

OUTCOME	INDICATOR [c]	1996-97 TARGETS	DATA SOURCE
C. A variety of courses are offered that had not been available.	9. Number of courses, by type, provided to participants— by educational institutions, and for various customer groups such as K-12 students (by grade/group), adult learners, educators, parents, and community members. (I)		School records, project records
	10. Number of schools offering Star Schools' high school credit courses, college preparatory courses, or advanced placement courses that had not been available previously. (I)		Project records, school records
	11. Number of students enrolled in Star Schools' high school credit courses, college preparatory courses, or advanced placement courses that had not been available previously. (I)		School records
	12. Number and percentage of students who enrolled in vocational education courses offered through the Star Schools program. (I)		School records
OBJECTIVE #2: **IMPROVED INSTRUCTION THROUGH DISTANCE LEARNING TECHNOLOGIES**			
A. Changes in instructional methods that integrate distance learning technology into the curriculum and are based on student learning needs.	13. Number of school administrators reporting changes / improvement in classroom teaching and learning practices, related to distance learning. (I)[e]		Surveys of school administrators
	14. Number and percentage of teachers reporting changes / improvements in teaching strategies or classroom practice as a result of participating in Star Schools. (I)		Teacher surveys
	15. Number and percentage of teachers reporting that the learning materials for students were an effective means to adequately cover the subject topics. (I)		Teacher surveys
	16. Number and percentage of teachers reporting that they use distance learning technology regularly in classrooms. (I)		Teacher surveys
	17. Number of teachers adopting innovative practices modeled on the Star Schools program, by subject matter and grade. (I)		Teacher surveys, surveys of school administrators
	18. Approximate number of educational hours during which distance learning activities were used by teachers. (I)		Teacher surveys

OUTCOME	INDICATOR [c]	1996-97 TARGETS	DATA SOURCE
B. Quality of distance learning equipment and materials.	19. Number and percentage of schools reporting that the distance learning equipment was (a) available, (b) adequate for their purposes, and (c) working properly. (I)		Surveys of school administrators, teacher surveys
	20. Number and percentage of schools reporting that the distance learning materials they had received were (a) understandable, (b) adequate/complete, and (c) being used. (I)		Surveys of school administrators, teacher surveys
C. Teachers and other educators receive improved professional development using distance learning technology.	21. Number of teachers or other educators who completed a staff development program in which distance learning was a significant training component. (I)[c]		Project records, teacher surveys, school records
	22. Number of teacher or other educator professional development activities (courses, workshops, special broadcasts) offered via distance learning technologies, by type. (I)		School records, project records
	23. Number and percentage of teachers and other educators rating as satisfactory the (a) understandability, (b) relevance, (c) adequacy, and (d) effectiveness of professional development activities that use distance learning approaches. (I)		School administrator surveys teacher surveys
D. Longer term use of distance learning activities.	24. Number and percentage of teachers who continue to participate in Star Schools' staff development activities after the formal Star Schools project has ended. (I)		Follow-up surveys of teachers, school records
	25. Number of schools that continue to use distance learning services (courses, staff development) when project is no longer receiving Star Schools funding, by characteristics of activities that continue. (I)		Follow-up surveys of school/ district administrators
E. Schools and other educational organizations continue to value and use educational opportunities delivered through distance learning technologies.	26. Number of other educational institutions (such as vocational and higher education) that continue to participate in programs after Star Schools' funding ends. (I)		Surveys of other educational institutions
	27. Number and percentage of school districts throughout the United States using distance learning activities, whether or not they have participated in Star Schools, by district characteristic (e.g., urban, suburban, or rural; income level, etc.) (I)		Nationwide survey of school districts

BEST COPY AVAILABLE

OUTCOME	INDICATOR [c]	1996-97 TARGETS	DATA SOURCE
OBJECTIVE #3: IMPROVED STUDENT LEARNING AND EMPLOYABILITY			
A. Improved student achievement on indicators that can be attributed at least partially to the Star Schools Program.[c]	28. Percentage of students who report that the distance learning activities added significantly to the quality of the course or subject, by various student characteristics, such as age, gender, ethnicity, etc. (I/E)[b]		Student surveys
	29. Percentage of students who report that they were able to apply the information learned via distance learning activities, by various student characteristics such as age, gender, and ethnicity. (E)		Student surveys
	30. Percentage of students reporting increased interest in school because of distance learning activities in their classes. (E)		Student surveys
	31. Number and percentage of students who found distance learning courses to be better than courses not using distance learning methods or activities. (I/E)		Student surveys
	32. Percentage of teachers reporting (a) increased learning by their students (b) improved student attendance (c) increased interest in subject area (d) improved critical thinking and problem solving, attributable at least in part to the use of distance learning activities in their classes. (E)		Teacher surveys

OUTCOME	INDICATOR [c]	1996-97 TARGETS	DATA SOURCE
	33. Percentage of students whose test scores improved significantly in courses in which distance learning technologies had been introduced and were a significant part of the instruction. (E)		School records
B. Change in employability.	34. Number of non-college bound students who are employed within (a) six months or (b) one year of completing high school and who report that participation in the program was a contributing factor in being selected for employment. (E)[h]		Ex-student surveys, school records

Notes:

a. The information displayed in this matrix was developed with staff members of the Star Schools Program at the U.S. Department of Education, with assistance from project directors and evaluators contracted by the program office between 1994 and 1996.

b. Major telecommunications technologies used to deliver distance learning include broadcasting, satellite, cable, telephone microwave, wireless, and computer networking. (Council of Chief State School Officers Distance Learning Report, 1995)

c. The data for all these indicators, except for indicators #27, #28, and perhaps #34, would be obtained from school systems participating in Star Schools projects. The primary data gathering is likely to be from active projects so that the longer term outcomes can be identified.

d. The Star Schools Program will need to decide whether the "program" refers to distance learning programs funded through Star Schools or more generally to all distance learning programs. This issue is relevant to multiple indicators in this matrix.

e. I = Intermediate Outcome Indicator

f. These three targets are illustrative only and not based on any official choices. Targets for each indicator should be established before the start of each year. Whenever percentage increases are specified as a target, the target should also indicate the baseline from which the percentage will be calculated.

g. Preferably these percentages would be broken out by amount of student exposure to distance learning technology, if a small number of exposure levels can be defined.

h. E = End Outcome Indicator

Appendix 4: Trained Observer Ratings

Step 4 on data sources and data collection procedures discussed the pros and cons of trained observer ratings and provided some examples of their applicability to education programs. This Appendix describes the various types of trained observer rating systems and summarizes the steps needed for developing these procedures.

Types of Trained Observer Rating Systems

Without a common yardstick to assess conditions, observers will likely come up with different answers. Trained observers who use systematic rating scales reduce the subjectivity in condition assessment. Rating scales, properly developed, provide trained observers with yardsticks against which to assess conditions.

Several forms of rating systems can be used by trained observers. These include:

- Written descriptions only,
- Photographic rating systems that use photographs as the rating scales,
- Other visual scales such as scales using drawings rather than photographs, or even videos,
- Combinations of the above.

Each of these is described briefly below.

Trained Observer Systems Using Only Written Descriptions. This is the simplest type of rating system. It depends solely on written descriptions of each grade used in the rating scale. These rating descriptions need to be quite specific about what constitutes each particular rating. This is necessary to maintain the reliability of the ratings.

Advantages of the written rating scales

- It is the "simplest" form of rating scale.
- It is a familiar procedure, used fairly often by some public agencies.
- It can also help local projects identify resource allocation needs, by specific locations where problems are present.

Disadvantages of the written rating systems

- The written information for each grade needs to be very specific in order to assure reasonable reliability, so that different raters will give approximately the same rating for a particular condition.
- Because of the need to clearly specify each grade on the rating scale, the trained observers are likely to have to spend a considerable amount of time developing the ratings.

Example: An abbreviated example of written guidelines for "building cleanliness" that a program may consider:

- Rating 1: **Clean.** Building is completely or almost completely clean; a maximum of three pieces of litter per floor are present.
- Rating 2: **Moderately Clean**. Building is largely clean; a few pieces of isolated litter/dirt are observable.
- Rating 3: **Moderately Dirty.** Some scattered litter/dirt is present on some floors.
- Rating 4: **Dirty.** Heavy litter/dirt is present in several locations throughout the building.

Systems Using Photographic Ratings. Photographic scales are one of the most useful forms of rating systems. They use pre-selected photos to represent the various grades on the rating scale. Each trained observer is given (and trained in the use of) a set of photos, perhaps four photos for each grade on the scale. This procedure is particularly useful for assessing the condition of facilities.

Other Visual Scales Visual rating scales can also use drawings or sketches to represent each grade on a rating scale that represents various outcome conditions. Exhibit 37 is an example of a rating scale using sketches to represent conditions of school buildings, in this case, the condition of school room walls. This was used by the New York City school system to track the physical condition of its schools. (The information was also used to help make decisions about building repairs.)

Major Steps in Implementing a Trained Observer Process

The following are steps needed to implement a trained observer process:

1. Decide what conditions should be rated.

2. Develop a rating scale for each condition to be rated. If possible, adapt an existing scale, modifying the rating system, photographs, and/or written guidelines as needed.

3. Determine which facilities/areas should be rated, when, and how frequently. Ratings can be applied to all or to selected areas or facilities. If you only have resources to rate *some* locations, choose them by using "random sampling" so that the locations chosen are representative.

4. Select and train observers. You can use program personnel, local project personnel, college or graduate school students, and volunteers. More technical ratings, however, will require persons with more "professional" training—to score open-ended tests or to obtain ratings of teacher use of new classroom teaching approaches.

5. Establish procedures for supervising observers, recording the data they collect, and transcribing and processing the data.

6. Conduct the ratings at the desired frequency.

7. Establish procedures for systematically checking the ratings of trained observers to assure quality control and reliability of the process. For example, the supervisor might check a small sample of the ratings done by each observer.

8. Develop and disseminate reports on the findings from each set of ratings. The reports will be more informative if they show the number and percentage of locations that fell into particular rating categories. Avoid reporting *only* the "average" scores, which can hide very important distribution information. Also, show changes from previous time periods.

Special Added Steps Needed for a Photographic Rating System

Here are the principal steps for developing your own scale in a photographic rating system:

- Take a large number of photographs in settings representative of the range of conditions to be rated. These photos should reflect the actual types of conditions that the program wants to assess.

- Select a panel of "judges" comprising persons with varied backgrounds who are not associated with the measurement activities. Select a set of familiar labels, each label representing a condition that the program expects to find such as: "clean," "moderately clean," "moderately dirty," and "dirty" (for observations of school facility cleanliness). Ask the judges to sort the photographs into groups that represent each condition.

- For each condition level, select these four or five photographs that most judges identified as representing that level.

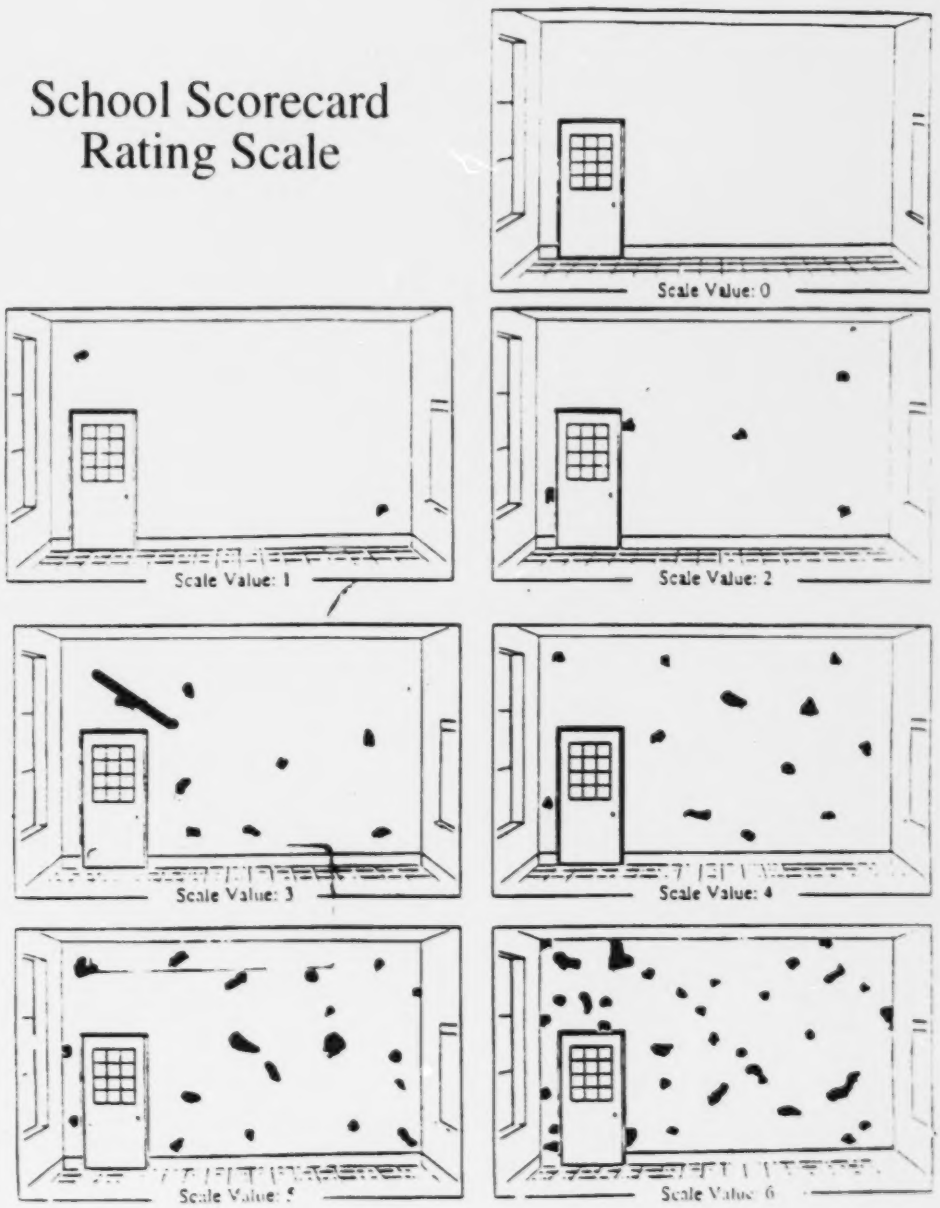
- If desired, develop written guidelines to accompany the photographs.

- Field test the scale with trained observers to determine if there is sufficient agreement among the observers on the ratings. Revise the procedures as needed.

- Develop the final scale. Package copies of the selected photographs in a "kit" for each trained observer.

See Reference section proceeding these appendices for citations on trained observer procedures.

Exhibit 37
Trained Observer Rating Scale:
"Condition of Classroom Walls"



☆ U.S. GOVERNMENT PRINTING OFFICE: 1997-517-199/83571